



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

Departamento de Fisiología y Biología Molecular y Celular

Análisis funcional de los determinantes genéticos y epigenéticos del contenido de vitamina E en tomate

**Tesis presentada para optar al título de Doctor de la
en el área CIENCIAS BIOLÓGICAS**

Estanislao Burgos

Director de Tesis: Dr. Fernando Carrari

Director asistente: Dr. Leandro Quadrana

Consejero de estudios: Dr. Norberto Iusem

Lugar de trabajo: Instituto de Biotecnología, CICVyA, INTA

Buenos Aires, 2018

Resumen

Análisis funcional de los determinantes genéticos y epigenéticos del contenido de vitamina E en tomate

El objetivo de esta tesis es dilucidar los determinantes genéticos de los contenidos de vitamina E (VTE) en tomate y su relación con el ambiente. Por un lado, se analizaron los contenidos de VTE en frutos de 146 entradas provenientes de distintos bancos de germoplasma y de 124 líneas recombinantes derivadas de un cruzamiento multiparental cultivados en ambientes contrastantes. Mediante estudios de asociación para el genoma completo (GWAS por sus siglas en inglés) y mapeo de QTL, se identificaron 60 *loci* potencialmente involucrados en la determinación de este carácter. Por otro lado, resultados previos del laboratorio indican que la expresión de los genes duplicados *VTE3(1)* y *VTE3(2)*, cuyos productos son responsables de los últimos pasos de la biosíntesis de VTE, son capaces de modular la acumulación de VTE en hojas y frutos de tomate. En particular, la expresión de *VTE3(1)* es regulada epigenéticamente debido a la metilación variable del ADN de la secuencia promotora. Con el objetivo de caracterizar las bases moleculares de esta regulación, como así también la co-regulación transcripcional de ambos parálogos, se generó y caracterizó molecularmente una población de 80 líneas recombinantes derivada del cruzamiento de líneas de introgresión portando los alelos silvestres (*S. pennellii*) y domesticados (*S. lycopersicum*) de dichos genes. Estos experimentos revelaron que la modulación transcripcional de *VTE3(1)* y *VTE3(2)*, mediada en parte por la variación en los niveles de metilación de sus promotores, es en efecto, un blanco importante para el mejoramiento de este carácter nutricional. En conjunto, esta tesis pone en evidencia la importancia de explorar tanto la variación intraespecífica como interespecífica para caracterizar las bases genéticas de fenotipos metabólicos y la necesidad de incluir la variación epigenética como un factor determinante de caracteres de interés agronómico.

Palabras Claves: Metabolismo secundario, Vitamina E, Epigenética, Parálogos, Tomate.

Abstract

Functional analysis of genetic and epigenetic determinants of Vitamin E content in tomato

The objective of this thesis is to elucidate the genetic determinants of the vitamin E (VTE) contents in tomato and their relationship with the environment. On one hand, VTE fruit's contents were analyzed for 150 entries from different germplasm banks as well as for 126 recombinant lines, which were derived from a multiparental crossing, cultivated in contrasting environments. By applying genome-wide association studies (GWAS) and QTL mapping, 60 *loci* were identified potentially involved in the determination of this character. On the other hand, previous results of our group indicate that expression of the duplicated genes *VTE3(1)* and *VTE3(2)*, whose products are responsible for the last steps of VTE biosynthesis, are capable of modulating the accumulation of VTE in leaves and tomato fruits. In particular, the expression of *VTE3(1)* is naturally and epigenetically regulated by variation in the level of DNA methylation of its promoter region. In order to characterize the molecular basis of this regulation, as well as the transcriptional co-regulation of both paralogs, a population of 80 recombinant lines derived from the crossing of introgression lines carrying the wild alleles (*S. pennellii*) and domesticated (*S. lycopersicum*) of both paralogs genes, were generated and molecularly characterized. These experiments revealed that the transcriptional modulation of *VTE3(1)* and *VTE3(2)*, partially regulated by the variation on the methylation levels of their promoters, is in fact an important target for the improvement of this nutritional character. Taken together this thesis highlights the importance of exploiting both intraspecific and interspecific variation to characterize the (epi)genetic basis of metabolic phenotypes and remarks the significance of include epigenetic variation as a determinant of agronomical important traits.

Key Words: Secondary Metabolism, Vitamin E, Epigenetics, Paralogs, tomato.

Agradecimientos

Quiero agradecer primero a las Instituciones que hicieron posible el desarrollo de esta Tesis. A la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires y al Instituto Nacional de Tecnología Agropecuaria (INTA Castelar) y en especial al Instituto de Biotecnología, quienes me permitieron seguir formándome como profesional.

A todos los que hicieron posible esta Tesis...

A mi Director, Dr. Fernando Carrari, por otorgarme la posibilidad de realizar este trabajo, por estimularme día a día, con su exigencia, su dedicación y compromiso. Por los buenos momentos vividos en INTA, en reuniones, congresos nacionales e internacionales, ensayos en Mendoza y por la posibilidad de realizar el intercambio en Francia. Por las enseñanzas que me brindó, por motivarme en todo momento y guiarme.

A mi Director asistente, Dr. Leandro Quadrana, por su dedicación al ayudarme y estar dispuesto a responder todas mis inquietudes (varias) desde el inicio de mi trabajo en el laboratorio.

A mis compañeros de grupo, Mariana Conte, Marina Insani, Guadalupe Dominguez, Carla Coluccio, Gabriel Litchestein, Luisa Bermudez, Andres Cernadas, Belén De Luca, Luis De Haro, por sus charlas, consejos, acompañamiento como equipo y buena predisposición en todo momento.

Gracias a mis compañeros de laboratorio por bancarme cuando tenía que moler muestras con la licuadora ruidosa. Me van a extrañar y yo a ustedes. Gracias a los chicos del invernáculo, por comprometerse con el trabajo, por cuidar las plantas, por ayudarme.

Gracias a Ramon Asis, Iris Peralta, Mathilde Causse, Christopher Sauvage y Elise por su ayuda y aportes a esta tesis.

A los grandes amigos que supe hacer en el INTA, Sebastián Moschen, Andrea Venturuzzi, Sole Lucero, Yamila Agrofolio, "La polaca", Gabi Conti, Ceci Rodriguez, Diego Zavallo, Juan Montechia y a todos los que forman parte de este gran grupo humano en el IB, sin ustedes esta tesis no hubiera sido posible. Gracias por estar, por acompañarme, por los tantos mates y cervezas que tomamos, por bancarme, gracias por hacer mas linda la estadía en el Instituto.

A todos los que me acompañaron a lo largo del doctorado... En especial a mis Padres, por estar conmigo siempre, apoyar todas mis decisiones y enseñarme a ser feliz. A mis hermanos Martin y Nacho, por ser cómplices incondicionales conmigo en la vida. A toda mi familia Abuti, Uri, Kity, Malo, Jorge, Laurita, Ale, Ceci, Vero, Vale, Conrado, Ariel, Ernestito, Viqui, Fran, Jaz y Rami que me acompañan y siguen todos mis pasos a través de la profesión. A Kari, Javier, Chavi, Nico, Sol, Cami, Lau, Corina, Jose, Pablo y Nacho con quienes comparto una hermosa amistad.

A mi gran amor Belén a quien conocí al otro lado de la mesada gracias por acompañarme en este camino

A todos, muchas gracias!

INDICE

Listado de abreviaturas.....	1
-------------------------------------	----------

CAPITULO I- Introducción.....	3
--------------------------------------	----------

I.1 El metabolismo vegetal; su importancia como fuente de moléculas nutracéuticas.....	3
I.1.1 El metabolismo de tococromanoles. Su importancia en la dieta humana y en las plantas.....	5
I.2 El fruto de tomate (<i>Solanum lycopersicum</i>) como modelo de estudio del metabolismo.....	11
I.3 Recursos genéticos y mejoramiento en tomate.....	13
I.3.1 Líneas de introgresión; poblaciones multiparentales; Mapeo genético; GWAS y mapeo de QTLs.....	15
I.3.2 Ventajas, limitaciones y complementación entre el mapeo de QTLs y GWAS.....	20
I.3.3 Criterio de evaluación de análisis GWAS.....	22
I.4. Bases genéticas del metabolismo.....	23
I.4.1 Interacción génica y alélica. Mecanismos epigenéticos asociados.....	23
I.5 Definición del problema, hipótesis y objetivos.....	27
I.5.1 Objetivo General.....	28
I.5.2 Objetivos específicos.....	29

CAPITULO II- Búsqueda de nuevos determinantes genéticos asociados a los contenidos de tococromanoles, aplicando dos estrategias complementarias: GWAS y mapeo de QTL.....	31
--	-----------

II.1 Introducción a las poblaciones estudiadas.....	31
II.1.1 Población de GWAS.....	32
II.1.2 Población multiparental -MAGIC-POP-.....	33
II.2. Resultados.....	36
II.2.1 Arquitectura del carácter estudiado.....	36
II.2.2 Detección de QTL.....	39
II.2.2.1 Por mapeo de asociación.....	40
II.2.2.2 Por mapeo de intervalo.....	41
II.2.3 Genes identificados por análisis de GWAS y mapeo por interval.....	52
II.2.4 Identificación de factores de transcripción asociados al metabolismo de tococromanoles.....	54
II.3. Discusión.....	58

CAPITULO III- Estudio de los mecanismos de interacción génica y alélica en la regulación del metabolismo de tococromanoles.....63

III.1 Introducción.....	63
III.2. Resultados.....	69
III.2.1 Obtención de una población F2 recombinante para los alelos silvestres y cultivados de los genes parálogos VTE3(1) y VTE3(2)	69
III.2.2 Fenotipado y análisis de expresión y metilación del ADN de los parentales de las líneas F2 recombinantes, desafiados en dos ambientes contrastantes.....	70
III.2.3. Caracterización fenotípica y molecular de individuos de la población F2.	76
III.2.4 Análisis de la población F2 y evaluación de la contribución de factores genéticos, epigenéticos y ambientales como así también sus interacciones en los niveles de tocoferol y expresión de VTE3(1) y VTE3(2)	81
III.3 Discusión.....	90

CAPITULO IV -Integración del conocimiento-Perspectivas.....93

CAPITULO V – Materiales y Métodos.....95

V.1 Descripción de los materiales y diseño experimental de una población de GWAS y 124 líneas recombinantes derivadas de un cruzamiento multiparental cultivados en ambientes contrastantes para la búsqueda de QTLs.....	95
V.1 .1 Materiales GWAS.....	95
V.1 .2 Materiales MAGIC POP.....	95
V.1 .3 Diseño experimental y condiciones de riego, ensayos GWAS y MAGIC-POP.....	96
V.1.4 Mapeo y datos genotípicos GWAS.....	97
V.1.5 Mapeo y datos genotípicos MAGIC POP.....	97
V.1.6 Identificación de asociaciones por GWAS a través de modelos MLM (Multitrait Lineal Mixed Model); cálculos de heredabilidad y porcentaje de la varianza explicada.....	98
V.1.7 Determinación de los IC (intervalos de confianza) que definen los QTL del GWAS.....	99
V.1 .8 Determinación QTLs de la F7 MAGICPOP.....	99
V.1.9 Búsqueda de Factores de transcripción.....	100
V.2.1 Descripción de la obtención de una población F2 proveniente de la cruce de dos líneas de introgresión.....	100
V.2.2 Cruzamientos.....	100
V.2.3 Desarrollo y aplicación de marcadores moleculares para el genotipado de la población recombinante F2 Cruza IL 3-1 x IL9.2.6.1.....	101

V.2.4 Diseño experimental y condiciones de la población recombinante F2 Cruza IL 3-1 x IL9.2.6.1.....	102
V.3 Medición de datos fenotípicos (contenido de tococromanoles) por HPLC.....	103
V.4.1 PCR en tiempo Real.....	103
V.4.2 Calculo de la expresión de los genes parálogos VTE3(1) y VTE3(2) en ambas condiciones (campo e invernáculo) y ambos tejidos (hoja y frutos) de la población F2 cruza IL3.1 X IL9.2.6.1 basada en la eficiencia de los primers alelo específicos.....	104
V.5 Análisis del nivel de metilación de la región promotora de los genes parálogos VTE3(1y2) mediante McrBC-PCRsc. Para la población F2: cruza IL3.1 X IL9.2.6.1, se evaluaron los 4 alelos, en dos condiciones (campo e invernáculo) y en dos tejidos (hoja fuente y frutos maduros)	105
CAPÍTULO VI - Referencias bibliográficas.....	107
Anexo.....	123

Índice de tablas y figuras

Figura 1. Esquema general de las vías del metabolismo primario y secundario...	4
Figura 2. Estructura de tococromanoles.....	6
Figura 3. Esquema de compartimentalización de un cloroplasto.....	7
Figura 4. Vista global de la fotosíntesis oxigenica, haciendo foco en la generación de ROS y la descodificación en la membrana tilacoide.....	9
Tabla 1. Genes clonados en tomate, su ubicación, fenotipos relacionados a calidad de fruto, hoja y arquitectura.	16
Figura 5. Bases genéticas cuantitativas.....	18
Tabla 2. GWAS y Mapeo de QTLs realizados para el estudio de caracteres de los tococromanoles en distintas especies hasta la fecha.	19
Figura 6. Efecto del fondo genético y la identificación de falsos positivos en Análisis de GWAS.....	22
Figura 7. Contextos genómicos de la metilación del DNA.....	23
Figura 8. Mecanismo descripto de RdDM.....	26
Figura 9. Estructura genética de la colección de GWAS.....	34
Figura 10. Construcción de la población MAGICPOP F7 a partir de 8 parentales.	35

Figura 11. Efecto fundador de cada uno de las 8 líneas parentales sobre la población de líneas recombinantes (F7).....	36
Figura 12. Variabilidad fenotípica para contenidos de tococromanoles en la población de de GWAS estudiada en ambas condiciones ambientales, control (C) y estrés hídrico (E).	37
Figura 13. Gráficos de cajas para contenidos de α-tocoferol para los distintos grupos genéticos dentro de la colección de GWAS en condiciones control (C) y de estrés (E).	38
Figura 14. Disección de la varianza fenotípica para los 6 caracteres medidos en el análisis de GWAS.....	39
Figura 15. Histogramas de distribuciones fenotípicas para los 6 caracteres medidos en las 124 líneas F7 de la MAGICPOP, posterior a la transformación \log_{10}.	39
Figura 16. Mapeo físico de los QTL determinados por GWAS y mapeo por intervalo (MAGICPOP).	42
Tabla 3. Número de genes en común entre el mapeo de QTL por intervalo obtenido a partir de la MAGICPOP y los 4 análisis de GWAS.....	43
Figura 17. Cantidad de genes comunes a los distintos análisis de GWAS con distintos grados de astringencia mediante un diagrama de Venn.	44
Tabla 4. Enriquecimiento de genes de la categoría Metabolismo Secundario (MS) en cada uno de los análisis aplicados (GWAS y mapeo por intervalos –QTL M-).....	46
Tabla 5. Enriquecimiento de genes relacionados al metabolismo de los tococromanoles para cada una de las estrategias de mapeo aplicadas.....	46
Tabla 6. Estimación de la heredabilidad y la varianza explicada por cada marcador asociado del análisis GWAS 2.....	47
Figura 18. Asociaciones identificadas para α-tocoferol en condiciones control asociadas a cada haplotipo identificadas a partir del análisis de GWAS 2.	49
Tabla 7. Medias fenotípicas asociadas a cada haplotipo por cada una de las asociaciones identificadas en el análisis de GWAS 2 para cada condición; marcador asociado; significancia; intervalos de confianza del QTL; extensión del mismo en Mpb y número de genes.....	50
Figura 19. Analisis de CM(2) a partir de datos de expresión, contenidos de tocoferoles y aporte parental de la MAGIC-POP.....	52

Figura 20. Análisis de VTE6 a partir de datos de expresión, contenidos de tocoferoles y aporte parental de la MAGIC-POP.....	54
Tabla 8. Factores de transcripción sobrerrepresentados en la vía de tococromanoles.....	55
Tabla 9. Factores de transcripción y motivos regulatorios en cis sobrerrepresentados en la vía de los tococromanoles e identificados en el análisis de GWAS 2.....	57
Figura 21. Ruta de biosíntesis de Vitamina E.....	59
Figura 22. Análisis de secuencias aminoácidas y nucleotídicas de los genes parálogos VTE3(1y2) en las especies S. lycopersicum y S. pennelli.....	64
Figura 23. Localización sub-celular de las proteínas VTE3(1) (A) y VTE3(2) (B) de S. lycopersicum en hojas de N. benthamiana.....	65
Figura 24. Árbol filogenético de VTE3.....	67
Figura 25. Genotipado de 8 plantas con distintas combinaciones de primers para los genes parálogos VTE3 de las 80 recombinantes F2 (cruzamiento IL9.2.6.1 x IL3.1) con dos marcadores moleculares ligados a VTE3(1) y VTE3(2).....	69
Figura 26. Contenidos de tocoferoles totales de frutos y hojas en condiciones contrastantes para los parentales de la F2; IL9.2.6.2, IL3.1 y para el padre recurrente cv M82.....	71
Figura 27. Gráfico de barras para niveles de expresión relativa de los ARNm de los genes VTE3(1) y VTE3(2) en condiciones contrastantes para los parentales de la F2, IL9.2.6.2, IL3.1 y el padre recurrente cv M82 en tejido de frutos y hojas.....	72
Figura 28. Gráfico de barras para niveles de metilación de la región promotora de los genes VTE3(1) y VTE3(2) en tejidos de fruto y hoja, de los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en condiciones contrastantes de invernáculo y campo.....	73
Figura 29. Relación de la acumulación de ($\alpha+\gamma$/ $\beta+\delta$) en frutos y hojas en condiciones contrastantes para los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en tejido de fruto y hoja.....	74
Figura 30. Esquema de obtención de los 9 genotipos recombinantes.....	77
Figura 31. Relación de la acumulación de ($\alpha+\gamma$/ $\beta+\delta$) en frutos en condiciones contrastantes de invernáculo y campo para los parentales de la F2, IL9.2.6.2,	

IL3.1 y M82 en comparación a sus genotipos reconstituyentes R1, R9 y R7 respectivamente.....	78
Figura 32. Niveles de expresión en frutos y hojas en condiciones contrastantes de invernáculo y campo para los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en comparación a sus genotipos reconstituyentes R1, R9 y R7 respectivamente.....	79
Figura 33. Gráfico de barras para niveles de metilación para la región promotora de los genes VTE3(1) y VTE3(2) en tejido de fruto y hoja, de los parentales de la F2, IL9.2.6.2, IL3.1 y M82 campo en comparación a sus genotipos reconstituyentes R1, R9 y R7 respectivamente, en condiciones contrastantes de invernáculo y campo.....	80
Figura 34. Gráficos de distribuciones de frecuencias para las distintas transformaciones ensayadas sobre los datos de δ-tocoferol en hojas de la F2 para los datos de δ-tocoferol de las 27 plantas clónales F2.....	82
Tabla 11 Resumen de la prueba de Shapiro-Wilks para todas las variables analizadas antes y después de aplicar la transformación BoxCox.....	83
Figura 35 Disección de la varianza fenotípica total para las distintas isoformas de tocoferol en frutos para los datos de toda la F2 cruza (IL9.2.6.1x IL3.1).....	85
Figura 36. Disección de la varianza fenotípica total para las distintas isoformas de tocoferol en hojas para los datos de toda la F2 cruza (IL9.2.6.1x IL3.1).....	86
Figura 37. Disección de la varianza asociada a la Expresión (Exp) y Metilación (Met) en tejidos de hojas (H) y frutos (F) a partir de modelos MLM para cada una de las variables teniendo en cuenta como factores : el ambiente, genotipo, GXE, metilación y planta para toda la población F2 cruza (IL9.2.6.1x IL3.1).	87
Figura 38. Gráfico de heatmap de correlaciones de Spearman entre variables fenotípicas de tocoferol, niveles relativos de expresión y metilación de los genes VTE3s así como de la sumatoria de la expresión de los cuatro alelos y su promedio de metilación para tejidos de hojas y frutos independientemente del ambiente.....	89
Figura 39. Fotografía del ensayo conducido en INRA Avignon en junio 2015.....	97
Tabla 12. Genotipado de la población Cruza IL 3-1 x IL9.2.6.1 para los genes VTE3(1y2).....	102

Tabla S1. Visión global del conjunto de materiales utilizados en esta tesis a través de los distintos capítulos.....	123
Tabla S2. Materiales GWAS: banco de germoplasma; origen geográfico; grupo genético; sub-grupo genético; referencias; color de fruto; contenidos de tococromanoles en frutos.....	124
Figura S1. Manhattan plots para las 32 asociaciones determinadas en el GWAS2.....	129
Figura S2. Box plots para las 32 asociaciones genotipo-fenotipo determinadas a partir del análisis de GWAS2.....	130
Tabla S3. Datos fenotípicos, contenidos de tococromanoles, de la población multiparental MAGIC-POP F7.....	135
Tabla S4. Asociaciones identificadas a través de 4 análisis de GWAS con distintos seteos para la determinación de las mismas.....	139
Tabla S5. Genes identificados a través de los distintos análisis de GWAS y el mapeo de QTLs reportados en la literatura, utilizados para la evaluación de enriquecimiento en la categoría genes relacionados a tococromanoles y para la búsqueda de FT sobrerrepresentados.....	142
Tabla S6. Ensayo de apilamiento de dos loci. F2 cruza (IL9-2-6-1 X IL3-1)	144
Tabla S7. Lista de primers utilizados.....	154

Listado de abreviaturas

DMPBQ: 2,3-dimetil 6-fitol 1,4-benzoquinona

GGDP: geranil geranil difosfato

HGA: ácido homogenístico

HGGT: homogentisato geranil geranil transferasa

LD: desequilibrio de ligamiento

MEP: ruta metabólica del metileritritol fosfato

MGGBQ: 2-metil 6-geranilgeranil benzoquinol

MP: metabolitos primarios

MPBQ: 2-metil 6-fitol 1,4-benzoquinona

MS: metabolitos secundarios

PDP: Fitol di fosfato

QTLs: caracteres cuantitativos asociados a una región genómica de un determinado locus del Inglés, *Quantitative Trait Loci*.

SK: Vía metabólica del Shikimato.

FT: factores de transcripción.

VTC: Vitamina C

VTE: Vitamina E

VTE1: enzima tocoferol Ciclase

VTE2: enzima homogentizato fitil Transferasa

VTE3(1): gen que codifica para la 2-metil-6-itolquinol metiltransferasa

VTE3(2): gen que codifica para 2-metil-6-fitolquinol metiltransferasa (de su sigla en inglés).

VTE5: enzima fitol quinasa

VTE6: enzima fitil fosfato quinasa

MAGIC POP: población avanzada de múltiples intercruces de su sigla en inglés).

MLMM: Modelo lineal mixto multi-locus (MLMM por sus siglas en inglés)

CAPITULO I- Introducción

1.1 El metabolismo vegetal; su importancia como fuente de moléculas nutraceuticas

Las plantas como organismos superiores se caracterizan por la complejidad de su red metabólica, aunque varían según la especie. Son organismos sésiles que han tenido que desarrollar diversas estrategias para su supervivencia. La síntesis de azúcares, aminoácidos, nucleótidos, lípidos, también conocidos como metabolitos primarios (MP), son los sustratos precursores para la síntesis de los metabolitos secundarios (MS), los cuales cumplen funciones ecológicas específicas, claves en el mantenimiento de la planta frente a animales herbívoros, infecciones microbianas o virales y la radiación UV. Estos incluyen desde las lectinas, los inhibidores de amilasa o de proteinasas y glucósidos o glucosinolatos cianogénicos, y a otras sustancias aditivas que atraen a los polinizadores, a modo de ejemplo.

La diversidad de MS producidos en el reino de las plantas es muy grande, existiendo cerca de 200.000 especies químicas estructuralmente distintas (Dixon y Strack 2003). Ciertos MS están restringidos a unas pocas especies de plantas, siendo un número muy pequeño redundante en un dado tejido y/o especie (Ferne 2007). Sin embargo, algunas especies han evolucionado a una diversidad particular presumiblemente por su adaptación a condiciones ambientales en las que se desarrollan (Li y col. 2015, Futuyma y Agrawal 2009). Esta es quizá la razón por la cual nuestro conocimiento de la estructura metabólica global es fragmentada para una gran cantidad de vías (Ferne y Tohge 2017).

Los MS se pueden dividir en tres grandes grupos: compuestos fenólicos, terpenoides isoprenoides y alcaloides glucosinolatos, que se producen a partir de diferentes vías del metabolismo primario: la glucólisis, el ciclo de Krebs, aminoácidos alifáticos, la ruta de las pentosas fosfato, vía del shikimato y aminoácidos aromáticos (figura 1).

Existe una estrecha relación entre los MP y los MS debido a que las vías del metabolismo secundario han evolucionado de forma parcial o completa a partir de otras vías del propio MS, pero también del primario. Es así, que muchos de los pasos en las vías son compartidos, tanto por sustratos, como enzimas, las cuales difieren en la especificidad y/o su localización celular y subcelular. Un buen ejemplo es el paso de generación de la cadena ramificada en leucina/metionina y la síntesis de glucosinolatos (Strassman y Ceci 1963).

La hipótesis del balance carbono/nitrógeno en plantas afirma que, bajo condiciones de disponibilidad limitada de nitrógeno, el MS se dirige a producir metabolitos ricos en carbono y viceversa, aunque existen excepciones (Coley y col. 1985). Se ha demostrado por ejemplo, como un incremento en la relación azúcar/aminoácidos resulta en una elevación en fenilpropanoides y en un decaimiento en los contenidos de alcaloides nicotínicos (Matt y col. 2002). Otro ejemplo de regulación entre MP y MS, es el caso del azufre, un componente esencial para plantas que cumple una serie de funciones en diferentes reacciones enzimáticas, en la regulación redox, metilación del ADN (dadores de grupos metilos), síntesis de poliaminas, como también es incorporado en una serie de MS. Cuando una planta crece en un ambiente deficiente en azufre, se produce el reciclado catabólico del azufre, siendo este un mecanismo crucial para responder a una serie de situaciones de estrés.

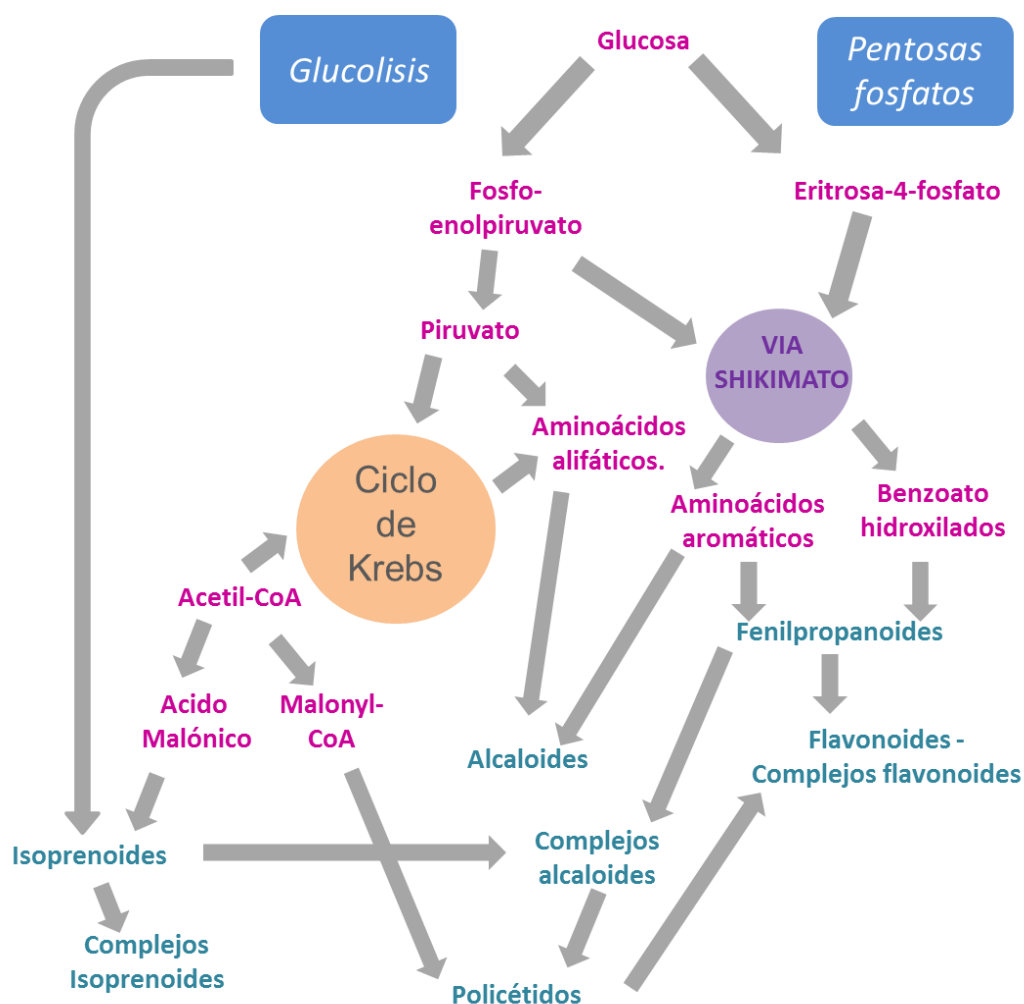


Figura 1. Esquema general de las vías del metabolismo primario y secundario. La producción de metabolitos secundarios (representados con color celeste) asociada a vías del metabolismo primario, ciclo de Krebs (en naranja) y central como la glicólisis (cajas celestes), la vía del shikimato (círculo violeta), la síntesis de aminoácidos aromáticos y alifáticos (fuxia). Adaptada de Aharoni y Galili (2011).

En los últimos años se han realizado importantes avances en la comprensión de la regulación de flujos metabólicos en las plantas. La expresión de genes que codifican para varios factores de transcripción (FT) asociados al metabolismo, ha demostrado que regula tanto la expresión de genes que codifican enzimas que participan del MS como a genes que codifican enzimas específicas involucradas en el MS relacionadas entre sí, indicando la existencia de redes altamente coordinadas entre MP y MS en las plantas (Maruyama y col. 2006, Malitsky y col. 2008, Sonderby y col. 2007).

Desde hace ya tiempo, el principal desafío para la biotecnología vegetal es aumentar, mediante ingeniería metabólica, los niveles de moléculas activas en plantas, ya que estos MS son útiles para múltiples propósitos, desde su empleo como tintas, saborizantes, fragancias, a estimulantes alucinógenos, insecticidas, agentes terapéuticos y nutricionales. La mejor forma para abordar este objetivo será de manera integrada entre los MP, MS y su regulación a distintos niveles.

1.1.1 El metabolismo de tococromanoles. Su importancia en la dieta humana y en las plantas

Los seres humanos necesitan ingerir nutrientes que les permitan desarrollar todos sus procesos vitales, siendo su fuente principal el consumo de vegetales. Entre estas moléculas esenciales que el organismo humano no tiene la capacidad intrínseca de sintetizar, y debe incorporar en la dieta, se incluyen las que tienen actividad de vitaminas, cuya función general es mantener la homeostasis del metabolismo celular promoviendo el correcto funcionamiento fisiológico. Las vitaminas pueden clasificarse por su solubilidad, siendo hidrosolubles las del complejo B y la vitamina C, mientras que la vitamina A (Retinol), D (Calciferol), K (Filoquinona) y E (Tococromanoles: tocoferoles y tocotrienoles) son liposolubles (Fitzpatrick y col. 2012). Hasta la fecha, 13 compuestos diferentes han sido clasificados como vitaminas. Cada una de estas moléculas tiene una función específica en las células vegetales, que puede diferir del mecanismo de acción que ejercen en células de mamíferos en el cual se las considera esenciales. A diferencia de los nutrientes principales, las vitaminas son requeridas en pequeñas cantidades (lo que las define como micronutrientes). En este sentido, la razón por la que algunos vegetales sintetizan (y acumulan) grandes cantidades de estas moléculas cuando su mecanismo de acción, si es el mismo que se describe para células de mamíferos, requeriría de pequeñas cantidades. Para el caso de los antioxidantes, como las vitaminas C (VTC) y E (VTE) la explicación surge de la explosión oxidativa que implican la fotosíntesis y la capacidad de lidiar con patógenos que colonizan poseen las células vegetales.

En particular, la VTE es el mayor antioxidante lipofílico incorporado en la dieta, y está constituida por un conjunto de moléculas denominadas tococromanoles (tocoferoles y tocotrienoles), sintetizadas mayoritariamente por organismos fotosintéticos. Durante mucho tiempo ha sido asumido que la biosíntesis de estos compuestos es exclusiva de las plantas, algas y algunas cianobacterias, todos organismos fotosintéticos. Sin embargo también se ha reportado en *Plasmodium falciparum* un organismo no fotosintético que causa la malaria, el cual sintetiza tococromanoles durante su estadio intraeritrocitario (Sussaman y col. 2011, Cassera y col. 2004). Los tocoferoles son moléculas anfipáticas constituidas por una cabeza polar cromanol, derivada del ácido homogentísico, y una cola lipofílica que deriva del fitol. Por su parte, los tocotrienoles son moléculas que difieren de los tocoferoles en la longitud e insaturación de la cola lipofílica (derivada del geranilgeranil difosfato), pero que encuentran amplia relación en su biosíntesis. Ambos grupos de moléculas se clasifican naturalmente en cuatro especies (α , β , γ y δ) metiladas diferencialmente en el anillo cromanol (Figura 2).

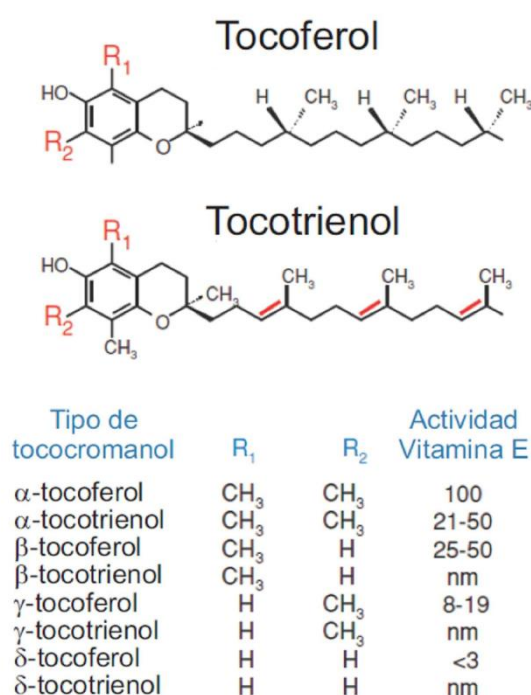


Figura 2. Estructura de tococromanoles. La figura muestra las moléculas de tocoferol y tocotrienol, indicando en rojo las diferencias entre las diferentes isoformas. En la parte inferior se especifica las sustituciones en el anillo cromanol para cada isoforma y su capacidad antioxidante relativa a α -tocoferol. “nm” refiere a que la capacidad no ha sido medida. Adaptado de DellaPenna y Pogson 2006.

La biosíntesis *de novo* de los tocoferoles se localiza en plástidos y algunas de las enzimas involucradas se encuentran en la membrana interna de

cloroplastos y cromoplastos, mientras que otras se ubican en estructuras membranosas asociadas a los tilacoides, llamadas plastoglóbulos (Vidi y col. 2006, Ytterberg y col. 2006, Miret y Munné-Bosch 2015) (Figura 3). En plantas superiores, los tococromanoles se han encontrado en mayor concentración en plástidos, coincidentemente con la localización de las enzimas de la vía central (Sun y col. 2009).

La confluencia de las vías del shikimato (SK) y del metileritritol fosfato (MEP) dan lugar al *core* de la síntesis de tococromanoles, o ruta central de su síntesis, la cual fue dilucidada por Soll y Schultz en 1979. Asimismo, el primer gen de esta vía fue clonado en *A. thaliana* en 1998, años siguientes integrando conocimientos obtenidos de organismos modelo como *A. thaliana* y *Synechocystis sp.*, permitió manipular diversos pasos de la biosíntesis (individuales o múltiples), a través del uso de mutantes y/o líneas transgénicas, modificando los contenidos y composición de estos compuestos tanto en hojas como en semillas de *A. thaliana* (Della Penna y Pogson, 2006) sugiriendo que existe cierta plasticidad en la producción de tococromanoles en plantas superiores (Falk y Munné-Bosch 2010, Mene-Safrane y Della Penna 2010).

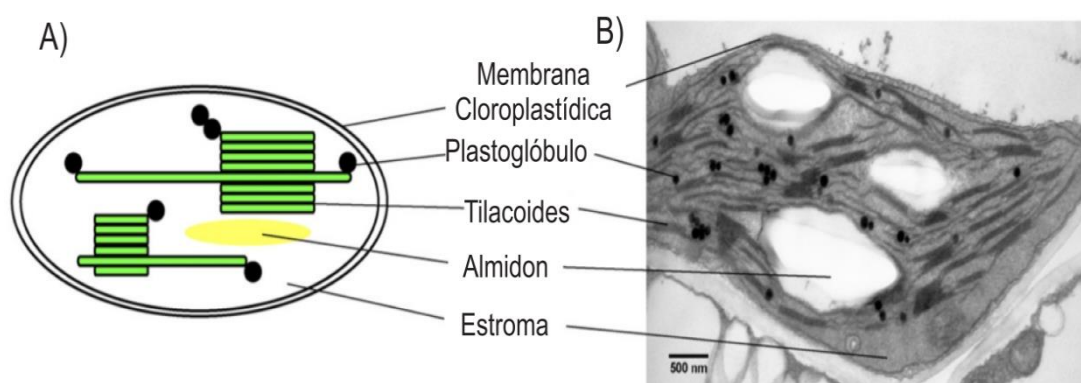


Figura 3. Esquema de compartimentalización de un cloroplasto. **A)** El esquema muestra el cloroplasto y su organización, delimitada por una doble membrana. El interior de los cloroplastos comprende el estroma acuoso, las membranas tilacoides (en verde), los gránulos de almidón (amarillos) y los plastoglóbulos (partículas negras), que son partículas lipídicas asociadas a la tilacoides. **B)** Esta imagen muestra en detalle un cloroplasto mediante microscopía electrónica de transferencia a partir de un corte ultrafino de tejido de hoja de *A. Thaliana*. Los plastoglóbulos son visibles luego de su tinción con tetroxido de osmio. En la parte inferior de la fotografía se muestra la escala que corresponde a 500 nm. Adaptada de Vidi 2007.

El primer paso del *core* ocurre a partir de PDP (fitil difosfato) y HGA (ácido homogentísico), los sustratos de, la homogentisato fitil transferasa (VTE2), la única enzima exclusiva de esta vía que produce 2-metil 6-fetil 1,4-benzoquinona (MPBQ). Posteriormente la tocoferol ciclaza (VTE1), cataliza la síntesis del anillo cromanol permitiendo la formación de δ -tocoferol. Paralelamente, la metilación de MPBQ por parte de VTE3 (2-metil 6-fetil metiltransferasa) forma 2,3-dimetil 6-fetil 1,4-benzoquinona (DMPBQ), que luego de la ciclación por

parte de VTE1, da como producto γ -tocoferol. Finalmente el último paso de la vía realizado por la enzima gama tocoferol metil transferasa (VTE4), culmina con la adición de un grupo metilo en la posición 6 del anillo cromanol, partiendo de δ -tocoferol y γ -tocopherol para producir β -tocopherol y α -tocopherol, respectivamente. El precursor PDP también proviene del reciclado de fitol como producto de la degradación de clorofila, por la acción de las enzimas fitol quinasa (VTE5) y la fitil fosfato quinasa (VTE6) (Valentin 2006, Vom Dorp y col. 2015).

Por otro lado, los tocotrienoles son sintetizados por la condensación de HGA con geranil geranil difosfato (GGDP), por la acción de la homogentisato geranil-geranil transferasa (HGGT). Como resultado, se produce 2-metil 6-geranil-geranil benzoquinol (MGGBQ), el cual es modificado de igual manera que en los tocoferoles por las enzimas VTE1, VTE3 y VTE4, para producir las cuatro formas: α -, β -, γ - y δ -tocotrienol.

Como se mencionó, la mayoría de las enzimas y actividades envueltas en la biosíntesis de tococromanoles están localizadas en la membrana de las tilacoides (Spicher y Kessler 2015). Sin embargo, el paso clave que es la ciclación por parte de VTE1 ocurre en los plastoglóbulos (Vidi y col. 2006) y el HGA se produce en el citosol (Wang y col. 2016), lo cual implica un tráfico de moléculas precursoras para que la biosíntesis de VTE ocurra. El tráfico debe ocurrir entre la membrana de los PG y el estroma (Spicher y Kessler, 2015), como también entre el citoplasma y el estroma (Pellaud y Mène Saffrané 2017). En este sentido, estudios previos de nuestro grupo, utilizando líneas de introgresión (ILs) de *S. pennelli* permitieron identificar un QTL (del inglés, *Quantitative Trait Loci*) para contenidos de tocoferol que mapea sobre el cromosoma 9. Este QTL co-localiza con un gen que codifica para una enzima del tipo SEC14-like. Miembros de esta familia están relacionados al transporte de tocoferol en mamíferos y en levaduras (Saito y col. 2007, Bankaitis y col. 2010).

Gran cantidad de reportes han demostrado el rol protector de los tococromanoles sobre los ácidos grasos polinsaturados frente a las especies reactivas de oxígeno (ROS) en un proceso conocido como peroxidación lipídica (revisado por Wolf 2005). Cuando esto ocurre el exceso de energía limita la capacidad de consumir todo el NADPH generado, como consecuencia se cierran los estomas para reducir la entrada de CO₂ y evitar así generar mayor cantidad de NADPH. La actividad antioxidante de los tococromanoles resulta de su reacción química directa con compuestos reactivos tales como radicales lipídicos, especies reactivas de oxígeno y de nitrógeno (Kamal-Eldin y Appelqvist 1996, Munné-Bosch y Alegre 2002, Schneider 2005) (figura 4). El mecanismo antioxidante de estos compuestos implica la transferencia de un átomo de hidrógeno fenólico del anillo cromanol a un radical peroxil lipídico

(ROO.), resultando en la producción de un hidroperóxido lipídico y un radical tococromanol. El radical tococromanol es más estable debido a la deslocalización del electrón desapareado y por lo tanto reduce su reactividad. Sin embargo, puede convertirse nuevamente en su tococromanol correspondiente por la reacción con otro reductor (por ejemplo, ascorbato o ubiquinol) (Kobayashi y DellaPenna 2008). Además, los tocoferoles, junto con los carotenos (Car) y el plastocromanol-8 (PC8), capturan oxígenos en su forma singlete ($^1\text{O}_2$), producidos durante la fotosíntesis por el fotosistema II, (Miret y Munné-Bosch 2015, Kruk y col. 2005, Kruk y Trebst 2008).

Más allá de su función antioxidante, los tococromanos también poseen un rol ecofisiológico muy importante en las plantas, el cual se extiende desde la

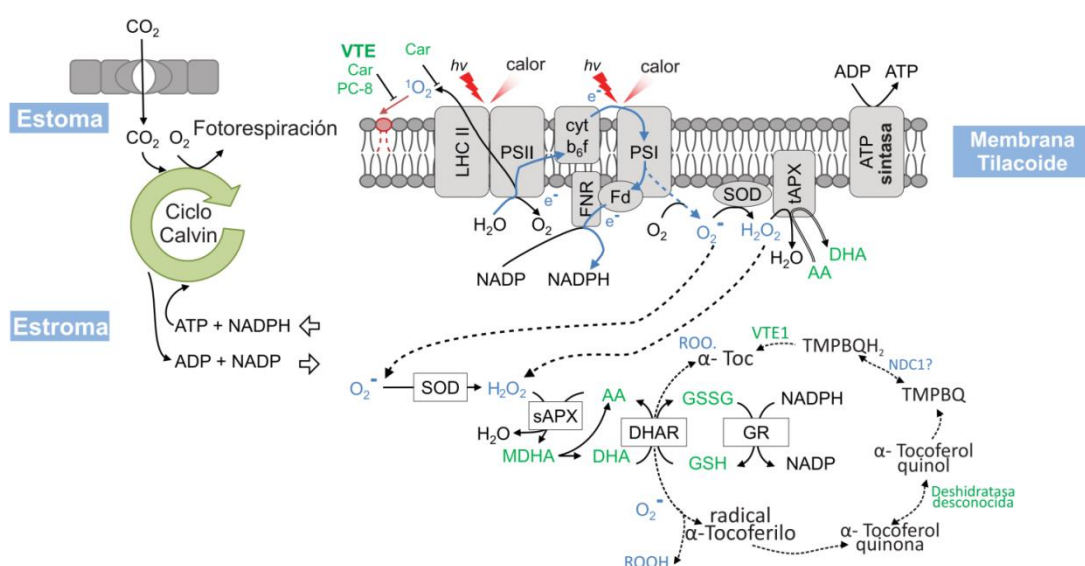


Figura 4. Esquema de la fotosíntesis oxigenica, haciendo foco en la generación de ROS y la descodificación en la membrana tilacoide. Todo el proceso dentro de los cloroplastos. Car (carotenos, Cyt B6f, citocromo b6f, DHAR, dehidroascorbato reductasa, fd, ferredoxina, ferredoxin-NADP⁺ oxidoreductasa, GR, glutatión reductasa, LHCII, *light harvesting complex II*, PSI, fotosistema II, sAPX ascorbato peroxidasa del estroma, SOD superóxido reductasa, tAPX ascorbato peroxidasa de las tilacoides. Ascorbato y glutatión tienen roles antioxidantes. MDHA (monodehidroascorbato) dismuta a DHA (dehidroascorbato) y AA (ácido L-ascorbico). Enzimáticamente convertir a AA consume NADPH. NDC1 NADPH-dependiente de quinona oxidoreductasa; VTE1, tocoferol cicliza TMPBQ, 2,3,5-trimetil-6-fetil-1,4-benzoquinona; TMPBQH2, 2,3,5-trimetil-6-fetil-1,4-benzoquinol. Ambos, los singletes de oxígeno ($^1\text{O}_2$) y radicales superóxido (O_2^-) y en especial el radical hidróxido, no representado en esta figura, pueden oxidar proteínas, lípidos y ácidos nucleicos si no son secuestrados por los mecanismos antioxidantes. Para simplificación solo el daño a las membranas lipídicas está representado. La peroxidación de ácidos grasos conduce a la formación de radicales peróxido lipídicos (ROO.) los cuales fácilmente se difunden en los sistemas biológicos. Estos reaccionan 1000 veces más rápido con el α -tocoferol (α -Toc) que con los ácidos grasos. El grupo hidroxilo del anillo cromanol puede donar su hidrógeno para expulsar los radicales peróxido lipídicos evitando su propagación. La presencia de otros antioxidantes como la Ascorbato AA es requerida para regenerar la capacidad antioxidante del α -tocoferol. Adaptado de (Miret y Munné-Bosch 2015)

germinación, la partición de fotoasimilados, la división celular, el crecimiento, la senescencia foliar, hasta las respuestas al estrés biótico y abiótico (Miret y Munne-Bosch 2015, Falk y Munné-Bosch 2010). Además, varios estudios han demostrado una estrecha interacción entre el metabolismo de tocoferoles y otras vías. Frutos de tomate que sobre-expresan la fitoeno sintasa (PSY), una enzima clave en la biosíntesis de carotenoides, muestran mayores niveles de tocoferol (Fraser y col. 2007). Por otro lado, los contenidos de tococromanoles se ven afectados cuando la vía de post-corismato es manipulada. Plantas transgénicas de *A. thaliana* sobreexpresando el gen bacteriano que sintetiza la enzima bifuncional corismato mutasa / prefenato deshidratasa (CM/PDT), muestran niveles significativamente más altos de fenilalanina, así como de γ -tocoferol y γ -tocotrienol, además de otros metabolitos secundarios (Tzin y col. 2009). Más recientemente, se ha encontrado estrecha relación con la degradación de clorofilas en el reciclado de su precursor PDP (Diepenbrock y col. 2017, Wang y col. 2018, Valentin y col. 2006, Vom Dorp y col. 2015). También se ha estudiado la variación de VTE mediada por señales endógenas y ambientales. Un claro ejemplo es el factor de respuesta a la luz DETIOLATED1, el cual resulta en frutos de tomate con mayores niveles de antioxidantes, entre ellos carotenoides, flavonoides y tocoferol (Davuluri y col. 2005, Enfissi y col. 2010).

Como se mencionó previamente, estos compuestos tienen actividad de VTE en mamíferos. Entre las funciones descritas para estas moléculas se destacan su influencia en la actividad enzimática, las cascadas de señalización y la regulación génica. Sin embargo los mecanismos moleculares *in vivo* aún no han sido dilucidados (Azzi 2007, Traber y Atkinson 2007, Briglius-Flohé 2009). En tanto, la forma α es considerada la de mayor actividad de VTE dado a que se almacena y transporta de manera eficaz en células de mamíferos (Traber y Sies, 1996). A pesar de que es inusual la deficiencia en VTE en la población occidental (Dror y Allen, 2011), modelos experimentales animales evidenciaron síntomas específicos vinculados a la insuficiencia de esta vitamina, como por ejemplo resorción fetal, distrofia muscular y encefalomalacia. Esta situación específica se ha encontrado en personas que no pueden metabolizar grasas (patología digestiva), en individuos con desórdenes en el metabolismo de lípidos (patología genética) y en niños prematuros con un peso corporal bajo. Más aún, ensayos realizados bajo dietas pobres en VTE se han asociado a una mayor predisposición a enfermedades como asma, cáncer, aterosclerosis y degeneración neuronal (McCary y col. 2011, Saldeen 2004, Terasawa y col. 2000, Yokota y col. 2001). Debido a que la dieta rica en grasas es la que proporciona la mayor absorción de VTE y en la actualidad hay una tendencia a reducir el consumo de grasas, esto conduciría a aumentar el consumo de tocoferol en la dieta de un adulto para así alcanzar los niveles recomendados de 15 mg/día (Bender y col. 2003). En este sentido, el consumo de vegetales

que aporten micronutrientes lipofílicos como la VTE es altamente recomendable. Proveer las bases del conocimiento que permitan entender los procesos regulatorios del metabolismo de los mismos es un pre-requisito para la explotación y consumo de dichos vegetales.

1.2 El fruto de tomate (*Solanum lycopersicum*) como modelo de estudio del metabolismo

La importancia económica de esta especie, sumado a las investigaciones básicas y aplicadas desarrolladas sobre este cultivo, dan lugar a que sea un modelo de estudio ideal. El tomate cultivado pertenece a la familia de las Solanaceas, que después de los cereales y las leguminosas, es la de mayor valor agregado en origen. Las mismas se diversificaron en América del Sur mucho antes de la separación de Gondwana, pruebas de esto lo dio la identificación de un fósil de 52.2 millones años de un fruto de tomatillo petrificado, encontrado en la Patagonia (Wilf y col. 2017). Esta hortaliza comprende más de 3.000 especies que se adaptan a los más diversos ambientes y dentro de esta familia, es la más cultivada mundialmente después de la papa. La batata, el aji, la berenjena y el tabaco son otros miembros de esta misma familia y constituyen otros modelos de estudio, como por ejemplo en el desarrollo de tubérculos (Ferne y col. 2001, Kolomiets y col. 2001) o en estudios de relaciones entre hospedantes y patógenos (Bogdanove y Martin 2000, Chung y Felton 2011, Gebhardt y Valkonen 2001, Martin y col. 1993, Pedley & Martin 2003, Rathjen y col. 1999, Rossi y col. 1998).

Las principales ventajas de la familia de las Solanaceas para ser utilizadas como modelo para el estudio de la diversidad y la adaptación ambiental, radican en su amplia variabilidad fenotípica y su conservación genómica y cromosómica demostrados a partir del análisis de cariotipo de diversas especies (Wang y col. 2008, Wu y Tanksley, 2010).

La sección *Lycopersicon* del género *Solanun* está formada por 1.500 especies, originarias en su mayoría del oeste de América del Sur (Ecuador y Perú) hasta el norte de Bolivia y Chile, a excepción de dos que son endémicas de las Islas Galapagos (Peralta y col. 2008). Las plantas ruderales de tomate cultivado se extendieron a otras partes de América, y en México fueron cultivadas y seleccionadas por los pueblos precolombinos (Jenkins, 1948). En la actualidad, el tomate está ampliamente distribuido a nivel mundial en cada uno de los continentes a excepción de la Antártida, ya que son cultivados o persisten en sus formas ruderales en ambientes alterados por la actividad del hombre.

El genoma de *S. lycopersicum* es diploide y relativamente pequeño (~900Mb). Esta es otra de las razones que lo posicionan como un modelo de investigación

alternativo al de *Arabidopsis thaliana* para el estudio de procesos de formación, maduración y metabolismo de frutos carnosos y climatéricos (Carrari y Fernie, 2006, Giovannoni y col. 2004). Además, esta especie es naturalmente autógama, las poblaciones sintéticas son estables y es capaz de cruzarse con otras especies del género (con *S. pimpinellifolium*, *S. habrochaites* y *S. pennellii*) de forma eficaz, sencilla y asistida. Estas características, muy importantes para su adopción como modelo de estudio, permitieron contar con amplios recursos genéticos y genómicos. Sumado a que posee protocolos de transformación génica de alta reproducibilidad (The Tomato Genome Consortium, 2012). Recientemente, se han publicado varios trabajos que dan cuenta de la posibilidad de aplicar edición génica en genes con diferentes mecanismos de acción (Brooks y col. 2014, Changtian y col. 2016, Soyk y col. 2018).

Dependiendo de las condiciones climáticas y de su crecimiento primario y secundario (por ejemplo la capacidad de desarrollar tallos secundarios), el tomate es una planta herbácea anual, bianual o perenne. La mayoría de las plantas poseen un crecimiento indeterminado, sin embargo algunas variedades son semideterminadas o determinadas, formando arbustos más ramificados y compacto. El tallo principal es simpodial, formando ejes laterales con hojas alternadas dispuestas en espiral de 1/3, en la que cada unidad finaliza con una inflorescencia.

Por su parte, el fruto es una baya compuesta por tejido epidérmico, un pericarpio grueso y la placenta gelatinosa que envuelve a las semillas. El desarrollo del fruto se divide en 4 etapas, de dos a tres semanas de duración, bien determinadas que involucran, I) el inicio de la floración hasta la antesis, II) el desarrollo del fruto caracterizado por divisiones celulares, III) la expansión celular sin divisiones, y finalmente, IV) la maduración determinada por modificaciones tanto en la composición metabólica (azúcares, aminoácidos y pigmentos) como en la composición de las paredes celulares (firmeza del fruto). Esta última etapa ha sido ampliamente estudiada tanto a nivel molecular como bioquímico, considerando su importancia en la cadena productiva en especial luego de la cosecha (Alba y col. 2005; Alexander & Grierson, 2002; Bapat y col. 2010; Carrari y col. 2006; de Jong y col. 2009). Está ampliamente descrito que este estadio de desarrollo se inicia una vez concluida la expansión celular, acompañado de un cambio abrupto en los perfiles metabólicos, un aumento de la hormona etileno y de la respiración (Carrari y Fernie, 2006). En tanto, se caracteriza por la des-diferenciación de cloroplastos y la concomitante degradación de clorofila, la re-diferenciación a cromoplastos con la acumulación de carotenos (licopeno y β -caroteno), dando lugar al color rojo característico de la maduración del fruto. Además, se producen modificaciones en la textura del fruto por variaciones en la turgencia celular y en la estructura de las paredes, acompañadas por cambios en la composición

de los principales metabolitos (azúcares, ácidos orgánicos) y compuestos volátiles, que determinan su calidad nutricional y organoléptica (Carrari y Fernie, 2006, Tieman 2006).

El fruto del tomate es una de las principales fuentes de tocoferoles para la dieta debido al elevado consumo en las poblaciones occidentales. Los datos publicados por el USDA en el año 2012 ubican al consumo de pasta de tomate en el 5to puesto entre 8.176 productos dietarios, por debajo de los cereales suplementarios (www.ars.usda.gov/ba/bhnrc/ndi). Además de VTE, el fruto del tomate aporta cantidades considerables de ácido ascórbico (VTC), licopenos, carotenos (pro-VTA) y compuestos fenólicos entre los principales antioxidantes. Existen numerosos estudios que describen las bases moleculares de la producción de estos compuestos en plantas superiores, aunque la mayoría hacen referencia a la VTC (Conklin y col. 2000, Jain y col. 1999) y licopenos (Auldrige y col. 2006, Simkim y col. 2004). En particular, han sido investigadas las diferentes vías de biosíntesis para la VTC mediante el clonado y mapeo de los genes involucrados (Zou y col. 2015). Además, otros autores han identificado la co-localización de algunos de estos genes con *loci* específicos de QTL para las variaciones en los contenidos de VTC en tres poblaciones diferentes de tomate silvestre (Stevens y col. 2006). Investigaciones realizadas en relación a los compuestos fenólicos aportados por el fruto del tomate señalan que los flavonoides y ácidos fenólicos, representados por quercetina y ácido clorogénico son los que se encuentran en mayor proporción (Slimestad y Verheul 2005). Estos antecedentes, sumado a la aplicación de estrategias de ingeniería metabólica, han permitido lograr frutos de tomate ricos en altos contenidos de estos compuestos (Giovinazzo y col. 2005, Muir y col. 2001, van der Rest y col. 2006). En contraste, el conocimiento desarrollado sobre la base de la regulación de las vías biosintéticas de tocoferoles en frutos de tomate es muy acotado (Almeida y col. 2011, Quadrana y col. 2014, Silvestre Lira y col. 2014). No obstante, los estudios realizados sobre el análisis de perfiles metabólicos en estos frutos indican la extensa variabilidad presentada entre las distintas especies y cultivares de tomate, proporcionando el conocimiento para el mejoramiento de estos caracteres (Schauer y col. 2005, Schauer y col. 2006; Quadrana y col. 2014).

En conjunto, estos antecedentes permiten proponer al cultivo de tomate como un modelo ideal para el estudio del metabolismo de tococromanoles.

1.3 Recursos genéticos y mejoramiento en tomate

Hace más de 200 años en Europa, se comenzaron a reproducir mejoras en el cultivar de tomate para satisfacer varias necesidades (Foolad 2007) haciendo foco en términos de rendimientos e índices de cosecha, actualmente los cultivares y/o híbridos, seleccionados por sus rendimientos redujeron de

manera considerable la base genética explotable del género *Solanum* (Bai y Lindhout 2007) y sólo un pequeño porcentaje de variables genéticas de la sección *Lycopersicon* fueron utilizadas para el mejoramiento. En consecuencia, no solo ha tenido un impacto negativo en la producción de compuestos de interés alimentario y farmacológico sino también en el estudio genético de las vías metabólicas de esta especie. En este sentido, un claro ejemplo es un trabajo que demuestra que el contenido de aminoácidos y pigmentos es entre 10 y 30 veces mayor en especies silvestres emparentadas respecto a los frutos de la especie domesticada (*S. lycopersicum*) (Schauer y col. 2005).

La mayoría de los caracteres implicados en la calidad del tomate se heredan cuantitativamente. Por esta razón, el tomate fue uno de los primeros cultivos a partir de los cuales se usaron marcadores moleculares para disectar la base genética de los rasgos cuantitativos o QTL a partir del trabajo pionero de Tanksley (1993). Debido al muy bajo polimorfismo revelado a nivel de especie, la mayoría de estos estudios se realizaron en progenies inter-específicas derivadas de cruces entre especies de tomate silvestre y el tomate cultivado (Swamy y Sarla 2008), potencialmente involucrados en la determinación de caracteres específicos que impactan directamente en su producción como ser peso del fruto, en los sólidos solubles totales o en el contenido de licopeno, entre otros, en *S. peruvianum* (Fulton y col. 1997), *S. pennelli* (Schauer y col. 2006; Kamenetzky y col. 2010) y *S. parviflorum* (Fulton y col. 2000). A partir de allí, estas regiones genómicas particulares se convirtieron en objetivos interesantes para la mejora del rendimiento del tomate (Swamy y Sarla 2008).

Más tarde, en los años 80, la vida útil era el problema más importante para los tomates frescos y los principales avances se centraron en elucidar la etapa del desarrollo de la maduración, principalmente porque sus determinantes afectan directamente la vida útil. Diferentes estudios se concentraron en la identificación de los principales componentes implicados en la maduración y firmeza de la fruta, incluido el papel del etileno (Alba y col. 2005) y la enzima poligalacturonasa (Giovannoni 2001). Pocas mutantes de maduración se caracterizaron y utilizaron en líneas heterocigotas y en el desarrollo de cultivares que exhibieron una maduración retardada, una mayor vida útil y una mayor firmeza. Más recientemente, se caracterizaron y mapearon varios genes relacionados con la maduración (Foolad 2007). Ya en los años 90, el sabor se convirtió en el objetivo principal. Éste es incluso un rasgo más complejo ya que no está claramente diseccionado y comprendido, principalmente porque está determinado por una gran cantidad de factores genéticos y no genéticos, no todos identificados o bien caracterizados (Causse y col. 2002). Se ha postulado que la relación entre azúcares y ácidos orgánicos es el principal determinante del sabor del tomate (Bucheli y col. 1999). Además, varios compuestos volátiles aromáticos tienen importantes influencias sobre el sabor (Goff y Klee 2006). A pesar de algunos ejemplos que explican el esfuerzo para mejorar esta

propiedad de la fruta, esta tarea se basa principalmente en los incrementos de sólidos solubles y acidez obtenidos en algunos casos (Foolad 2007). Finalmente, en la última y actual etapa, el objetivo de la mejora se centra en la calidad nutricional de los frutos de esta especie (Bai y Lindhout 2007).

Hasta la fecha se han identificado gran cantidad de polimorfismos involucrados en el desarrollo y la composición del fruto asociados a la calidad. En la tabla 1 se enumeran los principales, que afectan directa o indirectamente la calidad nutricional. Algunos de estos, pueden inducir variaciones en el color o aspecto de la fruta. Otras mutaciones, afectan a la arquitectura de la planta, como *sp*, que controla el crecimiento determinado/indeterminado, afectando la calidad (Pnueli y col. 1998). Otro ejemplo es la mutación *rin* (gen involucrado en el proceso de maduración). Se estudiaron estos híbridos y se demostró que la presencia de la mutación *rin* reduce la preferencia del consumidor. Además los mismos son un 17% menos dulce, con menos aroma y frutos más harinosos, aunque la firmeza y el contenido de azúcar no varían. Estos resultados muestran la influencia negativa de la mutación *rin* en la preferencia del consumidor, pero también indicaron que cuando se transfiere a un híbrido con alto sabor, la influencia negativa de la mutación se reduce.

1.3.1 Líneas de introgresión y poblaciones multiparentales. Mapeo genético; GWAS y mapeo de QTL

El mapeo de QTL en poblaciones segregantes está basado en genotipar la progenie derivada de cruzamientos entre distintos genotipos y asociar marcadores moleculares al fenotipo de interés en la descendencia (Figura 5). Las poblaciones homocigotas requeridas para estos propósitos son obtenidas mediante autofecundaciones repetidas de una F1 como es el caso de las RILs (del inglés: recombinante inbred lines). Cada individuo de la población posee múltiples introgresiones y las líneas homocigotas se utilizan en el análisis de QTL. Otro tipo de población son las denominadas Líneas de Introgresión (o ILs, por sus siglas en inglés), las cuales son obtenidas por repetidos retrocruzamientos y extensivo genotipado y selección. La población de ILs consiste en un reducido número de fragmentos genómicos de un padre donante sobre el fondo genético de otro (el padre recurrente). Las plantas homocigotas para cada uno de los *locus* se utilizan en los análisis de QTL mediante el mapeo de marcadores moleculares que delimitan cada *locus* y datos fenotípicos para cada una de las ILs.

Para las RILs la posición del evento de recombinación es fija, por lo tanto puede incrementarse mediante el agregado de más líneas. En contraposición, las ILs pueden ser mejoradas minimizando el tamaño de la introgresión (Alseikh y col. 2013). Aunque superficialmente las ILs y RILs parecen similares, estas difieren totalmente en cuanto a su uso dado que la frecuencia

de recombinación es considerablemente mayor para el mismo tamaño de población en las RILs en comparación a las ILs.

Las ILs contienen una única introgresión por línea y son preferentemente indicadas para identificar QTL de pequeño efecto. Las RILs al poseer un mosaico de genomas linqueado son más apropiados para el estudio de interacciones epistáticas (Keurentjes y Fernie 2011). Este tipo de recursos genéticos son muy poderosos para exponer los determinantes genéticos subyacentes a los caracteres cuantitativos, ya que las variaciones fenotípicas que se observan en las ILs respecto del parental recurrente pueden ser atribuidas principalmente al segmento introgresado y/o a las interacciones de este con el resto del genoma.

Tabla 1. Genes clonados en tomate, su ubicación, fenotipos relacionados a calidad de fruto, hoja y arquitectura de la planta.

Nombre y denominación ^a	Cr ^b	Posición ^c	Fenotipo ^d	Referencia
<i>Colorless epidermis</i> (y)	1	71255600	Epidermis rosada	(Ballester y col. 2010)
<i>Lutescent-2</i> (L-2)	10	61858478	Alteración de cloroplastos	(Barry y col. 2012)
<i>green flesh</i> (Gr)	8	60582066	Frutos verdes	(Barry y col. 2008)
<i>potato leaf</i> (C)	6	42804036	Hojas simples	(Busch y col. 2011)
<i>phytoene synthase 1</i> (r)	3	8606749	Frutos amarillos	(Fray y Grienson 1993)
<i>cuticular water permeability 1</i> (cwp1)	4	63765366	Microfisuras /deshidratación del fruto	(Hovav y col. 2007)
<i>carotenoid isomerase</i> (t)	10	62006972	Frutos naranja	(Isaacson y col. 2007)
<i>compound inflorescence</i> (s)	2	36913957	Inflorescencia	(Lippman y col. 2008)
<i>colorless nonripening</i> (Cnr)	2	37323107	No madura	(Manning y col. 2006)
<i>Jointless</i> (j)	11	3640857	Ausencia de pedicelo	(Mao y col. 2000)
<i>Falsifora</i> (fa)	3	61162449	Inflorescencia frondosa	(Molinero Rosales y col. 1999)
<i>single flower truss</i> (sft)	3	30564833	<i>Single flower truss</i>	(Molinero Rosales y col. 2004)
<i>de etiolated 1</i> (hp-2)	1	46495644	Alta pigmentación	(mustili y col. 1999)
<i>self-pruning</i> (sp)	6	42361623	Hábito determinado	(Pnueli y col. 1998)
<i>uniform ripening</i> (u)	10	2293088	Incremento clorofila	(Powell y col. 2012)

<i>Delta (Del)</i>	12	2285372	Frutos naranja	(Ronen y col. 1999)
<i>Beta-carotene (B)</i>	6	42288127	Incremento Beta carotenos	(ronen y col. 2000)
<i>sucrose accumulator (sucr)</i>	3	47401871	Acumulación de sacarosa en frutos	(Sato y col. 1993)
<i>lateral suppresser (ls)</i>	7	64958148	Supresión de corola parcialmente estéril el macho	(Schumacher y col. 1999)
<i>Zeaxanthin epoxidase (hp-3)</i>	2	46947557	Alta pigmentación	(Thompson y col. 2000a)
<i>ripening inhibitor (rin)</i>	5	5217073	Nunca madura	(Vrebalov y col. 2002)
<i>Macrocalyx(mc)</i>	5	5214073	Largos sepalos	(Vrebalov y col. 2002)
<i>apophytochrome B2 (phyB2)</i>	5	62648223	Recepción luz roja	(weller y col. 2001)
<i>apophytochrome A (phyA)</i>	10	22854459	Rojo lejano	(weller y col. 2001)
<i>Never ripe (Nr)</i>	9	62631866	No madura	(Wilkinson et al. , 1995)
<i>Entire leaf (e)</i>	4	59354677	Reducida complejidad hoja	(Zhang y col. 2007)
^{a)} Principales polimorfismos que afectan caracteres relacionados con la calidad nutricional del fruto de tomate. ^{b)} En que cromosoma se ubica ^{c)} Posición física dentro del cromosoma respectivo. ^{d)} Fenotipo asociado a la mutación. Adaptada de Albert 2017				

En 1995, Eshed y Zamir presentaron el desarrollo y aplicación de una colección de 76 líneas de ILs, con segmentos únicos del genoma de la especie *Solanum pennelli* (entrada LA0716) introgresados en el fondo genético del cultivar M82 de *S. lycopersicum* (entrada LA3475) que actuó como padre recurrente de las 446 retrocruzas que realizaron (Eshed y Zamir 1995). Estas líneas, cuyos segmentos introgresados se encuentran delimitados por marcadores moleculares, cubren el genoma completo de la especie y exponen 107 regiones solapantes llamadas BINs. Con el fin de identificar *loci* para caracteres de interés, estas ILs han sido ampliamente caracterizadas mediante el mapeo de más de 2.000 QTL (revisado en Fernie y Tohge 2017). Más aún, el uso de este material facilita la incorporación de alelos salvajes a cultivares *elite* que permiten el mejoramiento del cultivo sin recurrir a cultivares transgénicos.

Nuevos *backcrosses* de la línea de introgresión dieron lugar a una población de BILs (del inglés, Backcrosses introgression lines) esta ha sido utilizada para identificar genes relacionados a los azúcares y otras diferencias entre metabolitos de *S. lycopersicum* y *S. pennelli*. Otras poblaciones de ILs que se crearon son: *S. lycopersicum* x *S. chmielewskii* (Ballester y col. 2016, Do y col. 2010), *S. lycopersicum* x *S. pimpinellifolium* (Barrantes y col. 2016). Inclusive

se han construido ILs entre distintos cultivares que tienen sus ventajas (Chai y col. 2006, Zanol y col 2009).

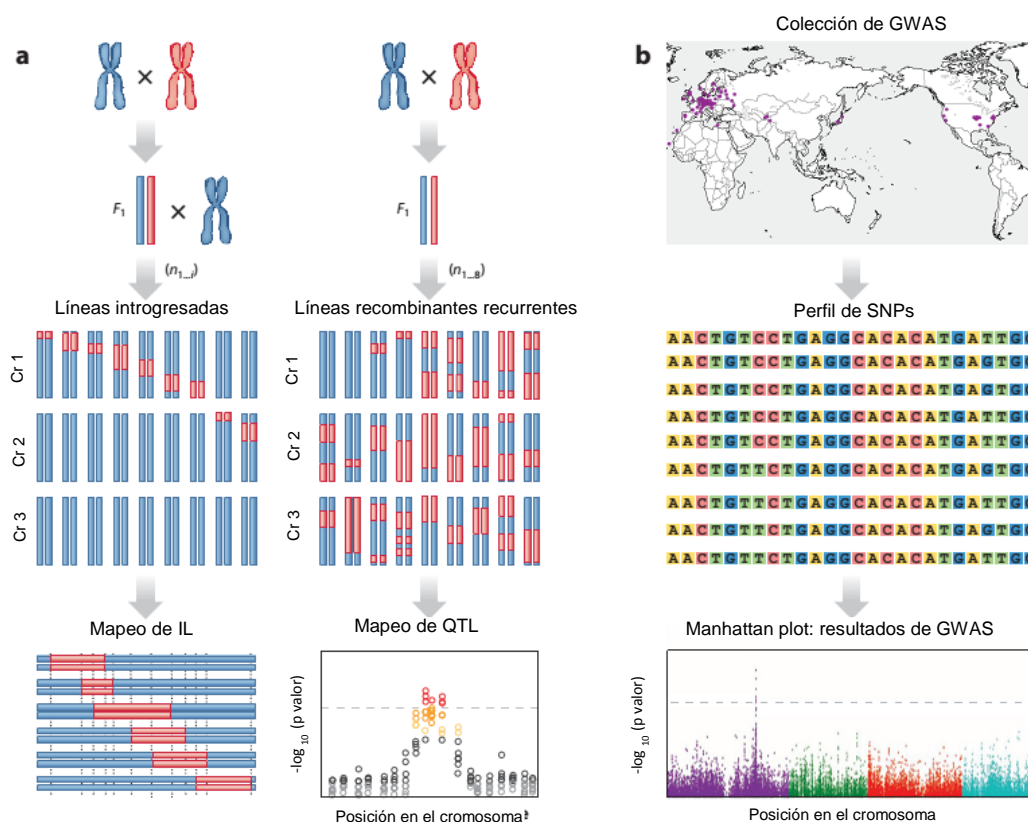


Figura 5. Bases genéticas cuantitativas a) Estructura de población de líneas de introgresión (ILs), líneas de retrocruzas recombinantes (RILs). Las ILs son creadas a partir del cruce entre dos líneas parentales y posterior cruce recurrente de la F1 con uno de ellos *backcrosses*. Luego las ILs homocigotas son aisladas mediante análisis de QTLs identificadas por mapeo de los datos fenotípicos. Las RILs son generadas por el cruce de una F1 al menos por 8 generaciones. Cada individuo en la población contiene múltiples introgresiones. Luego las líneas homocigotas son utilizadas para el análisis de QTLs. El hecho de que las RILs posean un mosaico de introgresiones permite el estudio de epistasis entre genes. **b)** La parte izquierda de la figura describe algunos de los pasos más importantes de un GWAS, mapeo de asociación de genoma amplio, generalmente se basan en utilizar la variabilidad natural asociadas al origen de los materiales utilizados, sobre ellos se mapean los distintos SNPs asociados al genoma para los distintos haplotipos y finalmente se realiza el análisis de asociaciones entre fenotipos y genotipos el cual está representado por el gráfico de Manhattan plot que posee en las ejes la posición de los distintos SNPs a lo largo de los distintos cromosomas del genoma y en las ordenadas se encuentra el grado de significancia de estas asociaciones $-\log$ del pvalor.

Más recientemente se ha comenzado a utilizar la vasta variabilidad genética presente en poblaciones naturales para identificar las bases genéticas de

números caracteres fenotípicos. En particular, el abordaje basado en GWAS (del inglés, Genome Wide Association Study) evalúa la asociación entre fenotipos para cada marcador polimórfico genotipado a lo largo de una extensa población de individuos genéticamente diversos (Figura 5b).

Los comienzos de los análisis de GWAS fueron orientados al desarrollo o uso de la genética médica (Zeng y col. 2015). En plantas, la implementación comenzó en *Arabidopsis* y dentro de los estudios de metabolismo asociados a defensa, en glucosinolato (Kliebenstein y col. 2001); enzimas del metabolismo primario y secundario (Mitchell-Olds y col. 1998), oligosacáridos solubles (Bentsink y col. 2000) y eficiencia en el uso del nitrógeno (Loudet 2003). La secuenciación más reciente del genoma del tomate, sumado a la liberación de matrices de genotipado han permitido genotipar simultáneamente miles de SNP a escala de genoma completo, ha allanado el camino para utilizar diseños de mapeo de QTLs más complejos, en los niveles inter e intraespecífico, para acotar la brecha entre fenotipo y genotipo. Trabajos que demuestran esto es el de Xu y col. (2013) en el cual utilizando un conjunto de 192 marcadores SNP genotificados en 188 accesiones, identificaron 2, 16 y 17 *loci* asociados a la acidez, los sólidos solubles y el contenido de azúcar, respectivamente. Otros reportes de QTL asociados a la calidad del fruto, como el pH, la acidez titulable, la SSC o la forma del fruto (Ruggieri y col. 2014, Sacco y col. 2015), lo que demuestra la viabilidad del mapeo de asociación para caracteres de calidad de frutos de tomate. En esta especie se han identificado 44 QTL relacionados a MS y 251 a caracteres asociados al metabolismo (Sauvage y col 2014, Tieman y col. 2017). En la tabla 2 pueden apreciarse todos los QTL reportados hasta la fecha en caracteres asociados a contenidos de tococromanoles en tomate y otras especies.

Tabla 2. GWAS y Mapeo de QTLs realizados para el estudio de caracteres de los tococromanoles en distintas especies hasta la fecha.

QTL ^a	Genes Candidatos ^b	Especie	Referencia	Metodología ^c
β-tocoferol	Tph1	Girasol	(Vera Ruiz y col. 2006)	Mapeo de QTL
α-tocoferol y total	No reportados	Maíz	(Wong y col. 2003)	Mapeo de QTL
31 QTL	HPPD,VTE3, VTE1,VTE4,VTE5	Maíz	(Chander y col. 2008)	Mapeo de QTL
α-γ-tocoferol y total	PDS y VTE3	Colsa	(Fritsche y col. 2012)	GWAS
16 QTL	Sec14, VTE3, VTE4, etc	Tomate	(Almeida y col. 2011)	Mapeo de QTL
51 QTL	VTE3, VTE2, SdS, CHL, otros	Maíz	(Diepenbrock y col.2017)	GWAS

73 QTL	COPII, CRY1, CYP59, dnaJ, LACS, o	Maíz	(Hong y col. 2018)	GWAS y Mapeo de QTL
Tocoferol total	No anotado	Tomate	(Sauvage y col. 2014)	GWAS
α -tocoferol	VTE4	Maíz	(Li y col. 2012)	GWAS
14 QTL	HPPD, VTE4, GST, Droug1	Maíz	(Lipka y col. 2013)	GWAS
14 QTL	VTE4, HPPD, IPP2, VTE5, TAT3, CMK	<i>A. Thaliana</i>	(Gilliland y col. 2006)	Mapeo de QTL
5 QTL	HPPD, VTE4, GST, SFH3, POD, SMT2	Girasol	(Haddadi y col. 2011)	Mapeo de QTL
94 QTL	No reportados	Colza	(Wang y col. 2012)	GWAS y Mapeo de QTL
5 QTL tocotrienoles	No reportados	Arroz	(Sookwong y col. 2009)	Mapeo de QTL
17 QTL	VTE4	Maiz	(Shutu y col. 2012)	Mapeo de QTL
11 QTL	HPPD	Maiz	Feng y col. 2013	Mapeo de QTL
18 QTL	No reportados	Soja	(Li y col. 2016)	Mapeo de QTL
^{a)} Tipo y/o número de QTL. ^{b)} gen candidato referenciado sobre el QTL. ^{c)} Metodología por la cual fue identificado el QTL				

1.3.2 Ventajas, limitaciones y complementación entre el mapeo por intervalos y GWAS

El mapeo de QTL por intervalos es una poderosa herramienta para identificar regiones del genoma que co-segregan para un carácter dado en poblaciones de familias F2 de líneas recombinantes. A pesar de su éxito, el mapeo de QTL por intervalo tiene dos limitaciones principales: i) la diversidad que segrega proveniente del cruzamiento de los parentales de la F2 es finita e innata a esas dos poblaciones únicamente. ii) como consecuencia, la cantidad de recombinaciones que ocurren durante la creación de las líneas recombinantes dan lugar a una limitante en la resolución de mapeo.

La primer limitante, diversidad alélica, puede ser incrementada hasta cierto punto intercruzando múltiples materiales/entradas (diversidad alélica) antes de establecer la población de RIL. Este es el caso de MAGIC (del inglés, Multi-parent Advanced Generation Inter-Cross) y en *Arabidopsis* AMPRIL (en inglés, multi-parent RIL) (Kover y col. 2009, Huang y Col. 2011). Estos nuevos tipos de

poblaciones que involucran varias líneas parentales han demostrado ser útiles en tomate para mapear QTL para rasgos de calidad en pequeños intervalos de confianza. Combinado con la secuenciación de las líneas parentales, permite el acceso directo a polimorfismos putativos bajo los QTL (Pascual y col. 2015). El segundo inconveniente de la capacidad de resolución de mapeo, puede ser mejorado mediante la generación de inter cruzamientos una vez establecida la población RIL, como por ejemplo en generaciones avanzadas de RILs (Balasubramanian y col. 2009).

Los análisis de GWAS y mapeo de QTL por intervalo son siempre complementarios cuando son conducidos juntos, mitigando de esta forma las limitaciones de uno en otro (Brachi y col 2010, Zhao y col. 2007). Por ejemplo los análisis de GWAS sirven como experimentos iniciales para explorar la arquitectura genética del carácter y a *posteriori* permiten elegir los padres para mapeo de QTL y luego validar los genes candidatos por mutagénesis, transgénesis y/o edición génica.

El poder del GWAS para identificar asociaciones verdaderas entre SNPs y caracteres depende de la varianza fenotípica de la población explicada por los SNPs. Esta, es determinada por la robustez de las diferencias fenotípicas de las variantes alélicas y sus frecuencias en la muestra. Es por esto que variantes alélicas raras de baja frecuencia son un problema para los análisis de GWAS (Asimit y col 2010, Gibsom y col. 2011). Estas variantes raras pueden estar asociadas a otras dentro del genoma y que sean no causales “sintéticas” las cuales pueden ser “arrastradas” por el desequilibrio de ligamiento (LD) (Gibson y col. 2010). Una forma de disectar la contribución de la heterogeneidad genética es incluyéndola como cofactor dentro de los modelos lineales mixtos (Segura y col. 2012). Es decir, el poder de detección de verdaderas asociaciones en un análisis de GWAS aumenta a medida que aumenta el número de accesiones analizadas, sin embargo esto ocurre, siempre que se este aumentando la variabilidad genética de la población, desbalances en la misma sobre ciertas frecuencias, de variantes no causales podrían llevar a falsos positivos (Figura 6). La elucidación sobre variables alélicas “raras” de baja frecuencia puede solucionarse usando análisis de QTL elevando la frecuencia de baja a intermedia a través de un esquema de cruzamientos que contemple este desbalance.

Hasta ahora la mayoría de los análisis desarrollados en *Arabidopsis* tienen solo unos cientos de individuos pero para cientos de caracteres. Resultados significativos pueden ser obtenidos con menos de 100 accesiones (Atwell y col. 2010). Esto se debe a que los caracteres considerados en estos trabajos son explicados por unos pocos *loci* que explican una gran porción de la varianza fenotípica. La situación es diferente en humanos donde típicamente un gran número de *loci* con pequeño efecto fueron encontrados para la mayoría de los

análisis, con lo cual, miles de individuos son requeridos en estos casos (Manolio 2009, Bopmer 2008) o por ejemplo *loci* con efectos intermedios como el color de ojos y piel han sido estudiados (Sulem y col. 2007, Beleza y col. 2013). La arquitectura genética en casos en los que son muchos los determinantes y de bajo efecto fueron observados en otros mamíferos (Flint y Mackay 2009) y en maíz (Tian y col. 2011). Esto muestra la existencia de diferentes arquitecturas genéticas de acuerdo al tipo de reproducción.

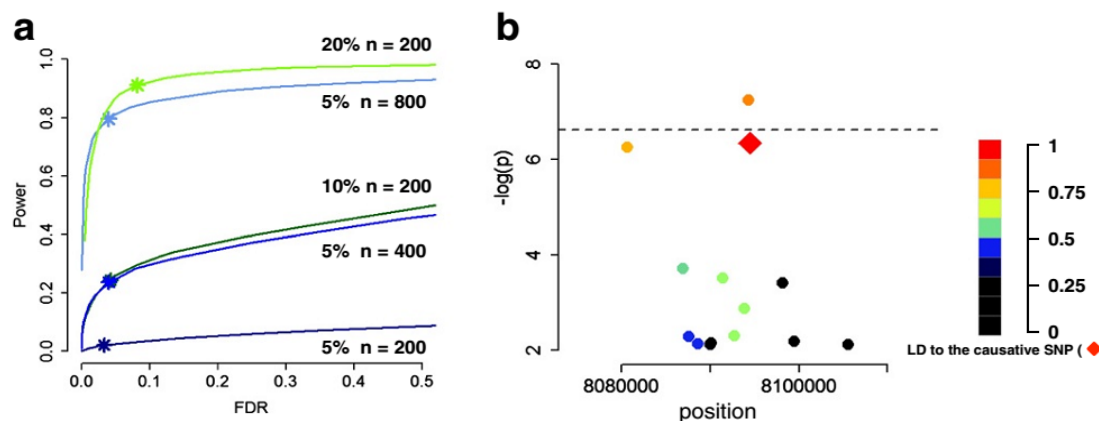


Figura 6. Efecto del fondo genético y la identificación de falsos positivos en análisis de GWAS a) Ejemplo en la identificación de falsos positivos para un fenotipo ideal. Simulación en la cual un marcador SNP explica 5%, 10%, 20% (con una heredabilidad cercana a 0,75%). Los asteriscos indican el poder par identificar verdaderos positivos y el FDR o del inglés *False positive rate*. **b)** Ejemplo de una situación particular en la cual un SNP causal, rombo rojo, no es significativo en el Manthattan plot. La escala de desequilibrio de ligamiento esta a la izquierda con una graduación de colores de negro (0) a rojo (1), la línea de corte responde al test de bonferroni para 220000 marcadores. Tomado de Korte y Farlow 2013.

1.3.3 Criterio de evaluación de analisis de GWAS

El criterio de evaluación de GWAS más básico e informativo de los abordajes es corregir por multiples pruebas (usualmente se utiliza el *test* de Bonferroni). La inspección de QQplots y Manthattan plots permiten evidenciar efectos de inflexión (Korte y Farlow 2013). Ambos abordajes, QQplots y Manthattan plots dan exploración global de los datos. Esto permite ver si hay muchas o pocas asociaciones significativas en relación a las esperadas. La principal limitación de esta corrección es asumir que cada SNP es independiente uno de otro. Las estructuras en las poblaciones violan este supuesto y hace que muchas falsas asociaciones persistan gracias al LD. El criterio más informativo es ver la proporción de falsos positivos y negativos simulada la cual tiene en cuenta el real LD de la población. Otra característica a evaluar es la robustez de los GWAS: calculando el enriquecimiento *a posteriori* evaluando genes anotados dentro de categorías funcionales (Kawakatsu y col. 2016). O *a priori* para evaluar el p-valor de corte para determinar si una asociación es significativa o no lo es, esta función debería ser creciente hasta determinado p-valor en el cual llega a un *plataou*, el cual presenta un máximo en el enriquecimiento de

genes identificados conocidos para la categoría para un p-valor umbral y en el cual no se determinan más asociaciones verdaderas y luego decrece.

1.4. Bases genéticas del metabolismo

1.4.1 Interacción génica y alélica. Mecanismos epigenéticos asociados

La epigenética podría definirse como la herencia de programas transcripcionales independientes de la secuencia de ADN. Existen una variedad de marcas epigenéticas; desde modificaciones covalentes del ADN como la metilación en citosinas, modificaciones postranscripcionales de histonas, incluyendo acetilación, metilación y ubiquitinación, variantes de histonas hasta ARNs no codificantes. En plantas, la metilación del ADN ocurre en contextos simétricos CG and CHG (H = C, A, o T) y en contextos asimétricos CHH (Law y Jacobsen, 2010). Los sitios de metilación CG se propagan durante la replicación del ADN mientras que la metilación de sitios CHG y CHH se mantienen tanto por propagación durante la replicación como a través de *self-reinforcing loops* que incluyen modificaciones de histonas y/o RNAi (Quadrana y Colot, 2016) (figura 7).

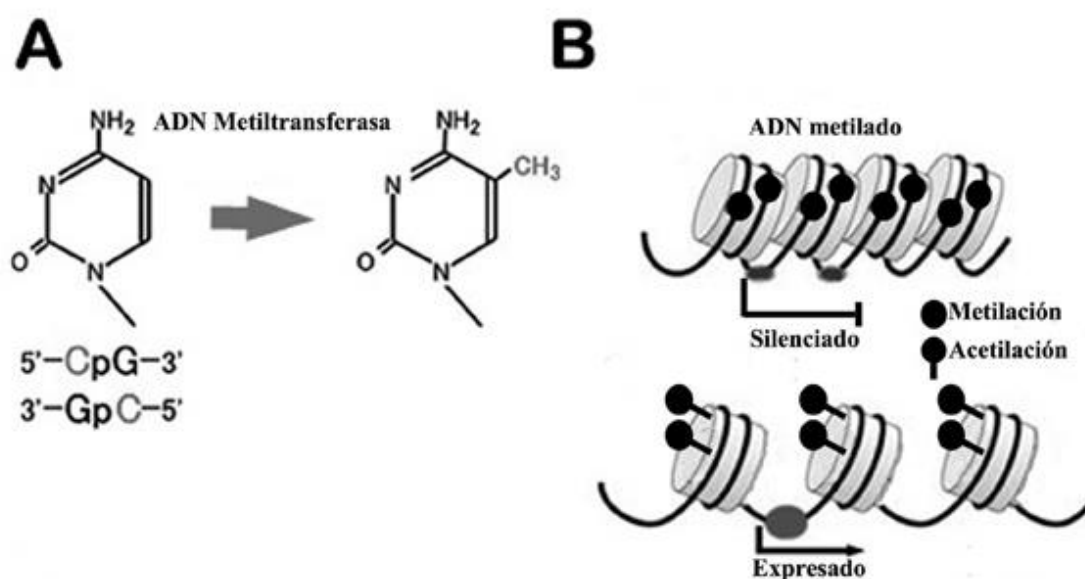


Figura 7. Contextos genómicos de la metilación del DNA A) Agregado del grupo metilo al C 5' de la citosina. B) Pasaje del estado de cromatina condensada y silenciada, por desatilación de histonas e hipermetilación del ADN, al estado descondensado y activado en el que la cromatina está desmetilada y acetilada. Tomado de Masuelli R y Marfil C 2011.

Variaciones en la metilación del ADN, llamadas "regiones diferencialmente metiladas (DMRs)", pueden generarse por diversas causas incluyendo: I) variaciones en la secuencia de ADN subyacente, II) mutaciones en factores en

trans involucrados en la mantención o remoción de la metilación (Pecinka y col. 2013), III) factores ambientales, entre ellos la temperatura y otros estreses (Downen y col. 2012, Dubin y col. 2015, Secco y col. 2015) o IV) por azar (Quadrana y Colot 2016)

Los estudios en plantas han proporcionado numerosas contribuciones fundamentales al campo de la epigenética. En este sentido, la primera evidencia de comportamientos epigenéticos no mendelianos se describe sobre la interacción entre dos alelos de un solo *locus* en determinados heterocigotas, que resulta en un cambio heredable de la expresión de un gen de uno de los alelos implicados, un fenómeno conocido como para-mutación observado en tomate y maíz (Arteaga-Vázquez y Chandler 2010). La impronta parental de genes individuales, es decir, la expresión de un único alelo de origen materno o paterno, que es un proceso cuya regulación es la base de múltiples trastornos genéticos en humanos (Zoghbi y Beaudet 2014), se observó por primera vez en maíz (Alleman y Doctor 2000). La repetida ocurrencia de individuos con simetría de flor alterada (de "peloria"), descrita por primera vez por Carl von Linneo en el siglo XVIII, es causada por una variación en el nivel de metilación entre dos alelos con idéntica secuencia de ADN, nombrándolos como variantes epialelica (Cubas y col. 1999). Estos epialelos pueden afectar patrones de desarrollo; en este sentido el gen *FWA* en *Arabidopsis* es un claro ejemplo donde el silenciamiento por la metilación del ADN de este gen retrasa la floración (Soppe y col. 2000). Además, un trabajo pionero sobre elementos transponibles en el maíz, realizado por Barbara McClintock y otros colaboradores en la década de 1940, reveló numerosos vínculos entre el comportamiento genético y la regulación epigenética (Lisch 2009).

Por otra parte y a medida que la tecnología transgénica en plantas como el tabaco, la petunia y *Arabidopsis* se convirtió en rutina a finales de la década de 1980, se hizo evidente que la expresión del transgén era impredecible (Wassenegger 2002, Matzke y Matzke 2004). Un caso sorprendente ocurrió cuando se intentó sobreexpresar el gen que codifica la chalcona sintasa (*CHS*), un regulador de la pigmentación de las flores en la petunia, e inesperadamente resultaron flores variegadas o incluso completamente blancas, en lugar de flores púrpuras más oscuras, fenómeno denominado "co-supresión". La falta de pigmentación en los sectores blancos se debió al silenciamiento génico coordinado tanto del transgén *CHS* como del gen *CHS* endógeno (Jorgensen y col. 2006). Esto condujo a constatar que el silenciamiento génico dependiente de homología de secuencia que ocurre cuando se introducen transgenes que llevan secuencias idénticas a genes ya presentes en el genoma, funciona como un "enhancer" de ARN mensajero denominado silenciamiento génico postranscripcional, PTGS. Junto con la represión de la transcripción por silenciamiento génico transcripcional (TGS) son las bases moleculares para el silenciamiento como herramienta biotecnológica. PTGS es una forma de

interferencia de ARN (ARNi), descubierta posteriormente en *Caenorhabditis elegans* y otros organismos.

A mediados de la década de 1990, se habían establecido vínculos entre el PTGS y la resistencia a los virus. Se demostró que los mecanismos epigenéticos que implican moléculas de ARN protegen naturalmente a las plantas de la replicación incontrolada de virus. Poco tiempo después se descubrió que estos mecanismos podrían explotarse para disminuir la expresión de genes de las plantas mediante la construcción de vectores virales que contenían secuencias de genes vegetales, aprovechando así la vía de silenciamiento génico inducido por el virus (VIGS, Baulcombe y Dean 2014; Senthil-Kumar y Mysore 2011). Otro mecanismo descubierto en plantas infectadas con viroides es la metilación del ADN mediada por ARN (RdDM) y otras marcas heterocromáticas en regiones de ADN homólogas (Wassenegger y col. 1994). Se demostró como se metilaban las citosinas que dirigen al promotor, dando como resultado el silenciamiento transcripcional dependiente de la homología del gen correspondiente (Mette y col. 2000) (figura 7).

Otra contribución destacada de la investigación en el área de la epigenética con plantas de tomate fue la aplicación de la secuenciación profunda para realizar análisis genómicos (Zhang y col. 2007), incluido el primer análisis de metilación genómica con resolución de una sola base (Cokus y col. 2008; Lister y col. 2008). Estos estudios requirieron el desarrollo de herramientas bioinformáticas que desde entonces se han aplicado ampliamente para la investigación. La metilación del ADN es una marca epigenética conservada importante para la integridad del genoma, el desarrollo y las respuestas ambientales en plantas y mamíferos. La desmetilación activa del ADN en plantas es iniciada por una familia de desmetilasas de ADN: 5-metilcitosina ADN glicosilasa/liasa bifuncional (Gong Z 2002; Gehring M 2006; Agius F y col. 2006; Morales-Ruiz y col. 2006). Se conoce que la maduración de frutos está asociada a cambios fisiológicos, bioquímicos y estructurales, como la acumulación de azúcares, compuestos volátiles aromáticos y pigmentos, y la hidrólisis de las paredes celulares (Klee y col. 2011). Estos cambios son inducidos por fitohormonas y factores de desarrollo. El etileno es uno de los factores más importantes que promueve la maduración, especialmente para las frutas climatéricas, como el tomate, que muestran un rápido aumento en el nivel de etileno durante la maduración (Klee y col. 2011). Como vimos anteriormente se han identificado varios reguladores de la maduración como: factores de transcripción (FT), inhibidores de la maduración (RIN), no maduran (NOR) y maduración incolora (CNR). Estos reguladores actúan río arriba en las vías de maduración dependientes de etileno como así también en aquellas vías independientes regulando la expresión de muchos genes relacionados con la maduración (Vrebalov y col. 2002; Manning y col. 2006; Giovannoni y col. 2007). Zhong y col (2013) mostraron evidencias acerca de como metilación del

ADN era otro participante importante en la maduración de frutos de tomate. Al aplicar un inhibidor de metilación de ADN se observó la regulación negativa de las desmetilasas identificadas, mediadas por ARNi, lo cual inhibió la maduración del fruto, corroborando la participación de marcas epigenéticas durante este proceso fisiológico (Zhong y col. 2013, Liu y col. 2015). Además, Zhang y col. (2013) adjudicaron la pérdida del sabor de frutos con la metilación de ciertas regiones genómicas portando loci que codifican para enzimas del metabolismo de compuestos volátiles. Hasta el momento, para *Solanum lycopersicum* hay identificadas cuatro desmetilasas de ADN: SIDML1 a 4; de acuerdo con la homología de secuencia con las de *Arabidopsis* ROS1 (AtROS1) (Liu y col. 2015, Cao y col. 2014).

Zhaobo Lang y col. (2017) compararon el metiloma de frutos de tomate con dos alelos mutantes de SIDML2, obtenidos mediante edición génica mediada por

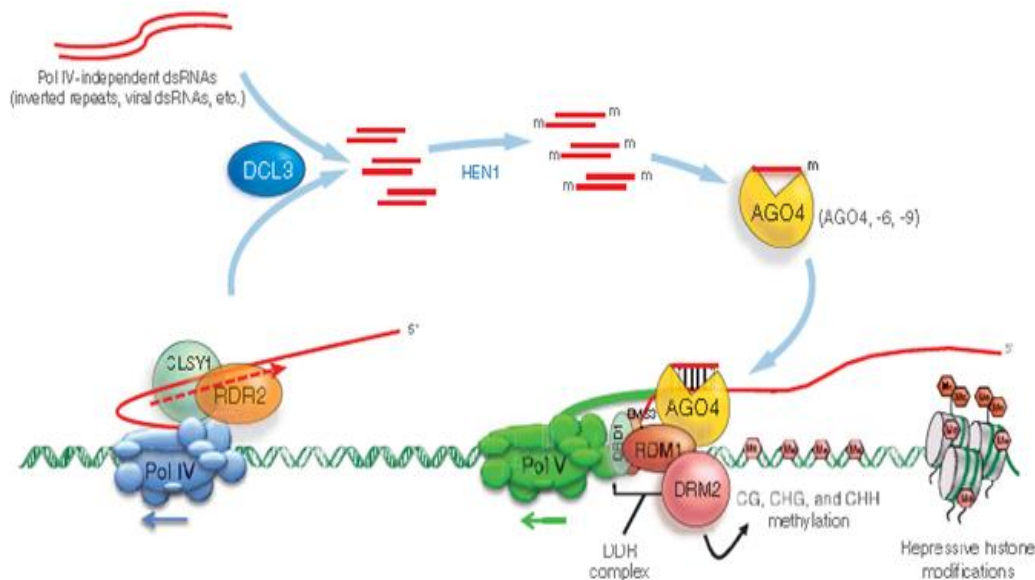


Figura 8. Mecanismo descrito de RdDM. Los retrotransposones, virus, transgenes o genes repetitivos están sujetos a silenciamiento por RdDM. Se cree que RNA Pol IV genera RNA monocatenarios (extremo inferior izquierdo) que sirven como plantillas para la RNA polimerasa dependiente de RNA, RDR2. Los dsRNAs resultantes se cortan en mediante DCL3 para generar dúplex de siRNA de 24 nt que luego se metilan por HEN1 y se cargan en AGO4, o sus miembros familiares estrechamente relacionados, AGO6 o AGO9. dsRNA derivado de transgenes repetidos invertidos o virus (arriba a la izquierda) puede eludir la necesidad de RNA Pol IV y RDR2. Los complejos AGO-siRNA se reclutan en sus sitios de acción uniéndose a transcripciones generadas por RNA Pol V, así como mediante interacciones físicas con el dominio carboxi-terminal de la subunidad más grande de RNA Pol V. En algunos loci, se cree que RNA Pol II sustituye RNA Pol V para la producción de transcripciones de andamios a las que se unen los complejos AGO-siRNA. El complejo DDR (DRD1, DMS3 y RDM1) permite la transcripción de RNA Pol IV. La subunidad RDM1 del complejo DDR también interactúa con AGO4 y la citosina metiltransferasa de novo, DRM2, lo que podría servir como un puente que recluta DRM2 a sitios de transcripción de RNA Pol V. Adaptado de Craig S. Pikaard y Ortrun Mittelsten Scheid 2017

CRISPR/Cas9, con respecto al control *wild type*. El análisis transcriptómico de los frutos de plantas *sldml2* sugirió que SIDML2 es necesaria para la activación de cientos de genes relacionados con la maduración, tales como PSY1, RIN y CNR y genes implicados en la síntesis de etileno y pigmentos, y la hidrólisis de la pared celular. Inesperadamente, el análisis también reveló que la desmetilación del ADN mediada por SIDML2 es necesaria para la represión de cientos de genes durante la maduración de los frutos. Estos genes reprimidos están implicados en procesos tales como la fotosíntesis y la síntesis de la pared celular, y su represión se correlaciona con la hipometilación del ADN en sus regiones promotoras.

1.5 Definición del problema, hipótesis y objetivos

La falta de inversiones en programas de mejoramiento genético sobre caracteres nutricionales complejos como el contenido de tococromanoles en cultivo de tomate se debe no solo a que estos han sido relegados frente a otros caracteres relacionados al rendimiento, sino también a la falta de conocimiento de su regulación a nivel molecular y su potencial implementación. Para que esto suceda, el primer paso es entender los mecanismos de regulación del metabolismo de estos compuestos y el rol de los factores genéticos, ambientales y epigenéticos que lo determinan.

Se ha estimado que el carácter presenta baja heredabilidad (h^2) en tomate (Schauer y col. 2008) y que la regulación de la biosíntesis está relacionada a proteínas de la maquinaria fotosintética que forman canales metabólicos en las membranas tilacoidales de *A. thaliana* (Tanaka y col. 2010).

Hasta el momento los únicos reportes centrados en el metabolismo de tocoferoles en tomate permitieron identificar y mapear 16 QTL como candidatos en la variación del contenido de estos compuestos: genes estructurales de las vías del MEP, SK y la vía central del tocoferol (Almeida y col. 2011).

Otro estudio mostró que la biosíntesis de tocoferol está regulada temporal y espacialmente manteniendo un balance constante del tocoferol total. Esto fue posible a través del análisis de ARNm de 47 genes involucrados en las vías biosintéticas mencionadas previamente en hojas y frutos en distintos estadios de desarrollo (Quadrana y col. 2013). Además, el análisis de las regiones promotoras de estos genes sugirió que los mismos son co-regulados por elementos comunes que actúan en *cis* (Quadrana y col. 2013).

Por otra parte el silenciamiento específico de dos genes estructurales de la vía *VTE3(1)* y *VTE4* en frutos (Quadrana y col. 2011, Quadrana y col. 2014), sugiere que la regulación puede explicarse, al menos en parte, a nivel de la actividad transcripcional de estos genes (Quadrana y col. 2013).

Más recientemente mediante el mapeo fino de un QTL mayor para los contenidos de α - y tocoferol total permitieron identificar al gen *VTE3(1)* como el responsable de este QTL. *VTE3(1)* codifica para una dimetil-fitolilquinol metil transferasa, la cual se encuentra en el cromosoma 9. Además este trabajo mostró que la regulación de su expresión diferencial entre el alelo silvestre y domesticado se debe a la presencia de un elemento transponible SINE, el cual se encuentra en el alelo domesticado y aquí se acumulan mayor cantidad de pequeños ARN (siARN) direccionando la metilación de su región promotora mediante un mecanismo de RdDM (Wassenegger y col. 1994) e inhibiendo su expresión, consecuentemente impactando sobre la acumulación de tocoferoles en los frutos de tomate (Quadrana y col. 2014).

Además, estos niveles revierten en respuesta a condiciones ambientales. Esto fue evidenciado al comparar los niveles de metilación que presenta la región proximal de este promotor en frutos maduros obtenidos de plantas cultivadas en condiciones controladas de invernáculos respecto de aquellos provenientes de plantas cultivadas en condiciones de campo (ensayo realizado en la Provincia de Mendoza, Argentina; 32° 50'S, 68° 52'O y a 900 msnm). En este mismo trabajo también se exploró la variabilidad epialélica de *VTE3(1)* en una población de *landraces* bajo un mismo ambiente, encontrando una plétora de epialelos los cuales presentan distintos niveles de expresión/metilación para la región del elemento SINE.

Por otro lado, un parálogo de aquel *VTE3(2)* mapeado en el cromosoma 3, se expresa preferencialmente en hojas (Quadrana 2013). A su vez, resultados disponibles en bases de datos (<http://ted.bti.cornell.edu/>), sugieren que *VTE3(2)* co-localiza con otro QTL para variaciones en los contenidos y composición de tocoferoles en tomate.

Aún cuando las evidencias resumidas más arriba sugieren que las variantes alélicas de *S. pennellii* de los genes *VTE(s)* podrían explicar las variaciones en los contenidos de VTE de los frutos, el efecto de otros *loci* involucrados en mecanismos aún no descriptos no puede descartarse.

Los problemas planteados guiaron a abordar el objetivo general de esta tesis:

1.5.1 Objetivo General:

Generar nuevo conocimiento y herramientas acerca de los mecanismos epi/genéticos que regulan los contenidos de vitamina E en frutos de tomates para su aplicación en programas de mejoramiento enfocados a aumentar su calidad nutricional.

A su vez, sobre la base de los antecedentes previamente descriptos se plantearon las siguientes hipótesis de trabajo que le dan sustento a esta Tesis:

Hipótesis 1: Dado que hasta el momento solo una pequeña parte de la variabilidad natural de los contenidos de tococromanoles puede ser explicada por los genes estructurales VTE, otros genes se encuentran adicionalmente involucrados en mecanismos aún no descritos que impactan en la acumulación de estos compuestos en las plantas de tomate. En este sentido y con el fin de explorar una mayor variabilidad genética e identificar nuevos *loci* candidatos asociados al carácter se abordan acá dos análisis complementarios de GWAS y mapeo de QTL por intervalos.

Hipótesis 2: La biosíntesis de tococromanoles es regulada diferencialmente a nivel de las enzimas involucradas y los genes *VTE3(1)* y *VTE3(2)* interactúan, determinando en parte los contenidos de estos compuestos en plantas de tomate. La comprensión de los mecanismos de interacción permitirá implementar mejores estrategias de mejoramiento de este carácter.

Para poner a prueba estas hipótesis se plantearon 3 objetivos específicos que se desarrollan en los capítulos II y III de esta tesis:

1.5.2 Objetivos específicos

- 1) Explorar otros determinantes genéticos de la variabilidad natural de tococromanoles mediante análisis un GWAS de una población de 146 genotipos y mapeo por asociación de una población de 124 líneas F7 multiparental (MAGIC-POP).
- 2) Analizar el efecto fenotípico sobre individuos F1 y F2 derivados de un cruzamiento entre una IL conteniendo el alelo silvestre de *VTE3(1)*(IL9-2-6-1) X otra conteniendo el alelo silvestre de *VTE3(2)* (IL3-1) en ambientes contrastantes.
- 3) Analizar los patrones de metilación y los niveles de expresión de los genes *VTE3(1y2)* en las líneas recombinantes productos del cruzamiento entre IL9-2-6-1 x IL3-1 en ambientes contrastantes.

CAPITULO II- Búsqueda de nuevos determinantes genéticos asociados a los contenidos de tococromanoles, aplicando dos estrategias complementarias: GWAS y mapeo de QTL

II.1 Introducción a las poblaciones estudiadas

La aplicación de la metodología de GWAS ha resultado en la exitosa identificación de numerosos genes involucrados en caracteres fenotípicos de base poligénica, varios de ellos validados a través de otras técnicas experimentales (Curtin y col. 2017). En particular, ha sido posible identificar genes asociados a contenidos de tococromanoles en frutos de tomate (Sauvage y col. 2014) y granos de maíz (Hong y col. 2018, Diepenbrock y col. 2017).

En este capítulo se presenta una integración de resultados obtenidos a partir de aplicar GWAS y mapeo de QTL con el fin de identificar determinantes genéticos de la variabilidad natural de tococromanoles en tomate. En particular, abordamos el estudio de una población de 146 genotipos especialmente seleccionadas para GWAS junto un mapeo de QTL a partir de una población de 124 líneas recombinantes F7 provenientes de un cruzamiento multiparental (MAGIC-POP) (ver en sección II.1.1 y II.1.2 para una explicación detallada de estas poblaciones como las condiciones de los ensayos y los datos genéticos de ambas poblaciones). Con este propósito, se analizaron los contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en frutos maduros provinieron de plantas crecidas en invernáculo. Asimismo, y con el fin de caracterizar posibles interacciones genotipo x ambiente, las mismas mediciones fueron realizadas en frutos de tomate de plantas cultivadas en condiciones de estrés hídrico.

La disección de caracteres cuantitativos es un desafío especialmente para especies de cultivo como el tomate. Si bien esta especie es considerada ampliamente diversa (Schauer y col. 2005), se dispone aún de una reducida capacidad para la medición y subsecuente explotación de la misma, sobre todo para caracteres relacionados al metabolismo de los frutos. Esto es más limitante aún cuando son utilizadas poblaciones con un alto grado de estructura genética (Blanca y col. 2015, Ranc y col. 2008). Para el caso del tomate, raramente se ha explotado la variación intraespecífica para realizar análisis como los que aquí se proponen (Tieman y col 2017). En lugar de eso, se han conducido estudios con progenies interespecíficas, a partir de cruzamientos entre especies relacionadas. Las dos poblaciones utilizadas en este trabajo permiten sortear estas limitaciones.

II.1.1 Población de GWAS

La colección completa de genotipos de la población de GWAS consiste en 210 entradas, de los cuales 146 son utilizados en esta Tesis (Tabla S2, suplementaria). Entre estas, 10 pertenecen a la especie *S. pimpinellifolium* (SP), 107 a la especie *S. lycopersicum* var. *cerasiforme* (SLC), 19 a un grupo genético mixto donde están incluidos materiales comerciales de tomates cherry de la colección del INRA (Francia) incluyendo materiales de *S. pimpinellifolium*, *Solanum lycopersicum* var. *cerasiforme* y *S. lycopersicum* var. *lycopersicum*. Para unas pocas entradas (10) no se dispone de información de sobre su origen. En términos del origen geográfico, un número mayoritario (40) fueron colectadas en Mesoamérica (México, Guatemala, Costa Rica, El Salvador, Honduras y Cuba), otro grupo importante (42 entradas) en el centro de origen (Perú, Ecuador, Colombia y Bolivia), un número similar de entradas (36) pertenecen a orígenes diversos y para otro grupo de 46 entradas no se cuenta con datos. Esta distribución geográfica es representativa de regiones de climas secos como las costas secas del Perú, y de valles húmedos como los de Ecuador. Estas entradas fueron incluidas en el panel con el interés de aumentar la diversidad genética de respuesta frente al déficit hídrico.

Entre estas entradas, 102 fueron previamente analizadas por Blanca y col. (2015), quien analizó el efecto de domesticación de 1.008 entradas ancestrales y contemporáneas a partir del 7.720 marcadores del tipo SNPs y específicamente sobre 6 *loci* que afectan el peso (*fw2.2* y *fw3.2*) y la forma de los frutos (*sun*, *ovate*, *fas* y *lc*). A partir del análisis de las frecuencias alélicas dentro de los distintos grupos genéticos y el lugar de procedencia, se propuso un proceso de domesticación en dos etapas; la primera durante la migración desde Ecuador del ancestro salvaje de las Islas Galápagos *S. pimpinellifolium* (SP) a Perú, donde se produjo la introducción de varias introgresiones dentro del cultivar *S. lycopersicum* var. *lycopersicum* (SLL) y luego en una segunda etapa en Mesoamérica. Además, este trabajo demuestra que *S. lycopersicum* var. *cerasiforme* (SLC) está posicionado filogenéticamente entre el ancestro salvaje SP y el cultivar SLL. Sin embargo, otros trabajos proponen que el primer proceso de selección ocurre durante la migración de Mesoamérica a Europa (Miller y col. 1990, Park y col. 2004).

Entre las entradas que conforman la colección de GWAS empleada en esta tesis, 72 fueron utilizadas para otro análisis de asociación publicado por Sauvage y col. (2014). En el mismo, se evaluaron 76 caracteres fenotípicos relacionados a MP y MS para un total de 163 genotipos correspondientes a SP (19 entradas), SLC (119 entradas) y SLL (28 entradas) utilizando 5.995 marcadores, durante dos campañas en distintos ambientes. Los resultados obtenidos permitieron identificar 44 asociaciones para 19 de los caracteres

estudiados. Las mismas se distribuyen sobre los 12 cromosomas de tomate, con al menos 2 asociaciones por cromosoma y relacionadas a contenidos de aminoácidos, azúcares, ascorbato (vitamina C), malato, citrato y contenido total de tocoferol. Esta última mapea sobre un gen del cromosoma 10 no anotado hasta el momento. Entre los genes candidatos encontrados a partir de estos análisis, se destaca un transportador de malato activado por aluminio (*Solyc06g072910*) el cual había sido caracterizado previamente (Martinoia y Rentsch 1995). Recientemente, Yan Yang y col (2017) confirmaron que un InDel de 3 pb en un dominio W-box de la región promotora impide la unión de un factor de transcripción represor del tipo WRKY42 y esto inactiva el transporte de malato que consecuentemente se acumula en las células.

Otro trabajo que demuestra la aplicabilidad de esta población es el reportado por Albert y col. (2016). Utilizando un modelo estadístico MLM se identificaron genes candidatos asociados a 27 caracteres de calidad de los frutos (pH, sólidos solubles totales; glucosa; fructosa; vitamina C; ácido málico; cítrico y firmeza del fruto) y de rendimiento del cultivo (tiempo hasta floración; diámetro de tallo; largo de la hoja; altura de la planta; número y peso del fruto; rendimiento y materia seca). A su vez, se registraron y analizaron estas mismas variables en condiciones de déficit hídrico. El estudio determinó 55 asociaciones que incrementan los contenidos de metabolitos, manteniendo el rendimiento en condiciones de estrés. Sumado a esto se identificaron 31 QTL únicos en la condición de déficit hídrico, otros 41 en la condición control y 44 QTL común a ambas condiciones. Interesantemente, se encontraron 25 QTL asociados a las interacciones con las condiciones ambientales. Once de estas regiones fueron reportadas también en los materiales de la población biparental de RIL provenientes de un cruzamiento entre los parentales de la MAGIC-POP (Pascual y col. 2015).

Los mismos 6.100 marcadores SNPs utilizados en el análisis de GWAS de esta tesis fueron utilizados previamente en el trabajo de Albert y col. (2016) donde además se estudió la estructura genética poblacional. La misma está compuesta por los tres grupos genéticos mencionados más arriba: SP, SLC y mixto y subgrupos de acuerdo a su lugar de procedencia (figura 9A y B), (más detalles de los marcadores se describen en la sección V.1.3 de Materiales y Métodos).

II.1.2 Población multiparental -MAGIC-POP-

La MAGIC-POP utilizada en este trabajo es la primera población multiparental de tomate que se encuentra públicamente disponible (Pascual y col. 2015), con un mapa de alta densidad cubre 758 Mb (84%) del genoma de tomate (900 Mb) para el estudio de la variabilidad intraespecífica. La misma permite el análisis

de nuevas combinaciones alélicas, además de la identificación de el/los polimorfismos causales de las variaciones aplicando mapeo de QTL. El estudio de la estructura de la MAGIC-POP a partir de los 1.345 marcadores del tipo

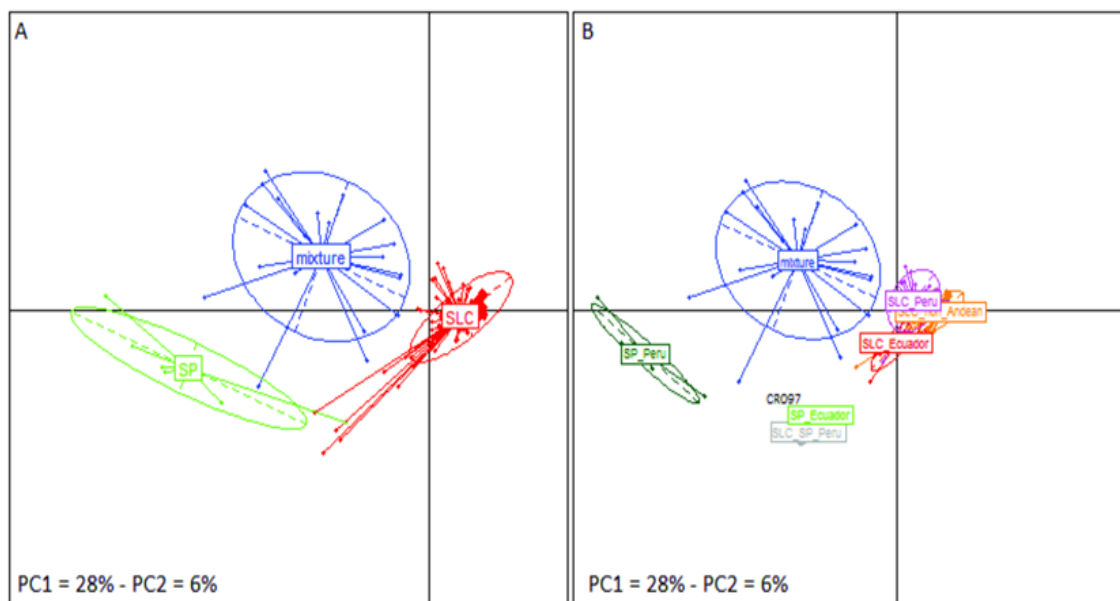


Figura 9. Estructura genética de la colección de GWAS. A) Análisis de coordenadas principales a partir de 6.100 marcadores SNPs utilizados en el análisis de GWAS presentado en esta tesis. En verde se distingue el grupo correspondiente a los individuos *Solanum pimpinellifolium* (SP); en rojo aquellos comprendidos dentro de *S. lycopersicum* var. *cerasiforme* y un tercer grupo genético fue determinado como “mixto” en color azul, el cual comprende más de una especie ver tabla suplementaria S2. **B)** El análisis de la población permite distinguir subgrupos de acuerdo a su lugar de origen (no andino SLC (violeta), SLC Perú (naranja), SLC Ecuador (rojo), SLC-SP Perú (gris claro), SP Perú (verde oscuro), SP Ecuador (verde claro) y la entrada CR097, la cual quedó indefinida. Las elipses de dispersión comprenden el 67% de los individuos para cada grupo. Adaptado de Albert y col. (2016).

SNPs incluidos en su mapa genético de acuerdo al test de Evanno y col. (2005) indicó ausencia de subgrupos. El primer trabajo en el cual se utiliza esta población (Pascual y col. 2015) identificó un total de 9 QTL para peso de fruto, 3 de estos estables a través de 2 ambientes. Posteriormente, Pascual y col. (2016) utilizaron estos materiales junto con una población de RIL biparental y un panel de GWAS con los mismos 163 materiales publicados anteriormente (Sauvage y col. 2014) para identificar un total de 71 QTL para peso de planta y racimo, tiempo a floración, peso fresco, firmeza, color, sólidos solubles totales y pH de los frutos. De estos, un total de 17 fueron compartidos entre la población de RIL y la MAGIC-POP, lo cual es consistente con el hecho de que la población bi-parental comparte dos de los parentales más distantes con la MAGIC-POP.

Esta población MAGIC-POP fue desarrollada a partir del cruzamiento entre 8 líneas parentales escogidas de un conjunto de 360 entradas (Ran y col. 2008, Figura 10) con el propósito de maximizar la diversidad genética. Para su construcción se realizaron 4 cruzamientos entre *S. lycopersicum* var. *esculetum* y *S. lycopersicum* var. *cerasiforme*, con frutos de menor y mayor tamaño, respectivamente. Los cuatro genotipos del grupo de *S. lycopersicum* var. *esculetum* elegidos fueron: Levovil (L1), Stupicke (L2), LA0147 (L3) y Ferum (L4). Del grupo de *S. lycopersicum* var. *cerasiforme* se escogieron los genotipos Cervil (C1), Criollo (C2), Plovdiv (C3) y LA1420 (C4). Posteriormente, los híbridos F1 Levovil x Cervil y Stupicke x Criollo fueron cruzados para obtener 120 plantas provenientes del doble cruzamiento de híbridos. En paralelo, los híbridos LA0147 x Plovdiv y Ferum x LA1420, también fueron cruzados para obtener otro grupo de 120 plantas doble híbridas. Finalmente, estos dos grupos de 120 plantas doble híbridas fueron cruzadas independientemente. Dos descendientes por cruzamiento fueron obtenidos produciendo 480 plantas, las

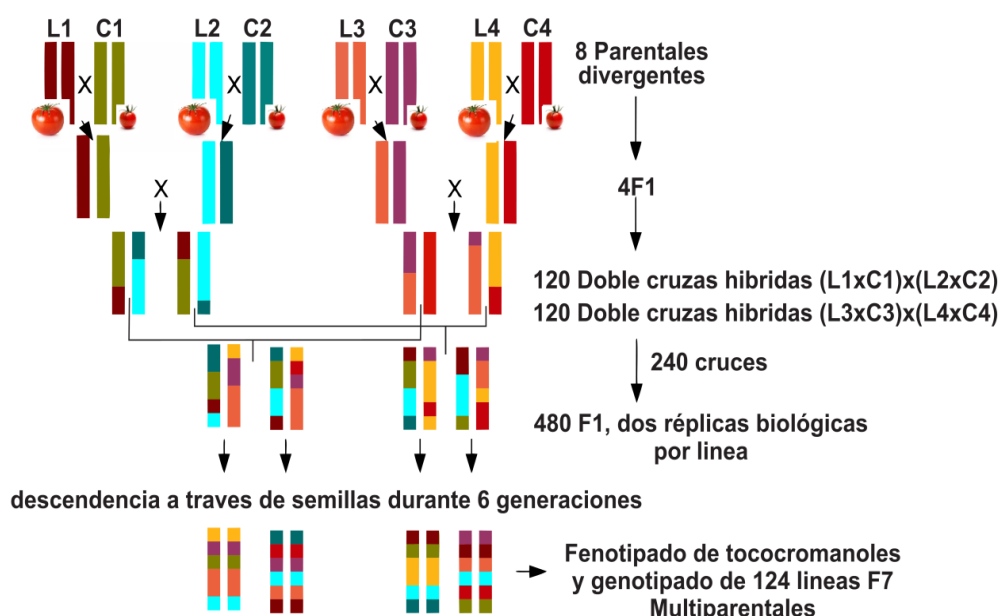


Figura 10. Construcción de la población MAGICPOP F7 a partir de 8 parentales. La población se construyó a partir de 8 líneas. Para su construcción se realizaron 4 cruzamientos entre *S. lycopersicum* var. *esculetum* y *S. lycopersicum* var. *cerasiforme*, con frutos de menor y mayor tamaño, respectivamente. Los cuatro genotipos del grupo de *S. lycopersicum* elegidos fueron: Levovil (L1), Stupicke (L2), LA0147 (L3) y Ferum (L4). Para el grupo de *S. lycopersicum* var. *cerasiforme* se escogieron los genotipos Cervil (C1), Criollo (C2), Plovdiv (C3) y LA1420 (C4). Posteriormente, los híbridos F1 Levovil x Cervil y Stupicke x Criollo fueron cruzados para obtener 120 plantas provenientes del doble cruce de híbridos F1. Simultáneamente, las líneas híbridas LA0147 x Plovdiv y Ferum x LA1420, se cruzaron para obtener otro grupo de 120 plantas doble híbridas F1. El paso final implicó el cruzamiento de ambos grupos de 120 plantas doble híbridas, produciendo 480 plantas F1, las cuales poseen regiones de cada uno de los 8 parentales fundadores en sus genomas. Estas plantas fueron propagadas a través de semillas durante 6 generaciones creando un conjunto de 397 líneas F7, de las cuales 124 fueron analizadas para el mapeo de QTL asociados a contenidos de tococromanoles de esta tesis

cuales poseen regiones para cada uno de los 8 parentales fundadores en sus genomas. Estas plantas fueron propagadas a través de semillas durante 6 generaciones de autofecundación creando un conjunto de 397 líneas F7, de las cuales 124 fueron analizadas para el mapeo de QTL asociados a los contenidos de tococromanoles en los frutos maduros en esta tesis. Es por esto que la población muestra un amplio rango de fenotipos y ausencia de estructura genética. El diseño utilizado para desarrollar esta población, integrando el efecto de los ocho genomas parentales (figura 10), permitió además un balance en las frecuencias alélicas a lo largo de todo el genoma (figura 11). La excepción en cuanto a la homogeneidad de frecuencias se presenta en los cromosomas 5 y 11 donde las 3 líneas parentales Stupicke, LA0147 y Levovil fueron más difíciles de discriminar.

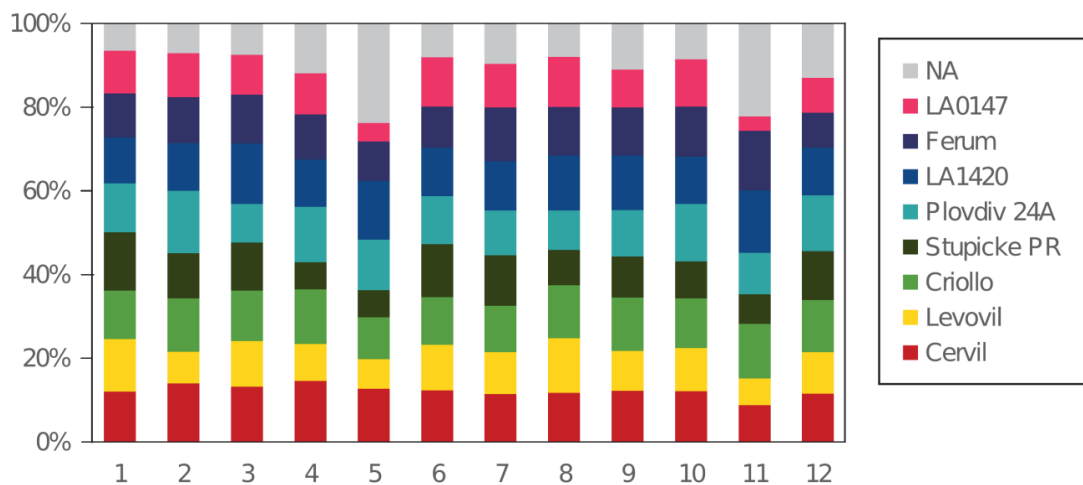


Figura 11. Efecto fundador de cada uno de las 8 líneas parentales sobre la población de líneas recombinantes (F7). En esta figura se puede observar un balance de frecuencias alélicas homogéneas para cada uno de los 12 cromosomas. En gris claro, está indicada la proporción del genoma a la cual no ha sido identificada la procedencia alélica. Tomado de Pascual y col. (2015).

II.2. Resultados

II.2.1 Arquitectura del carácter estudiado

Previamente a realizar el análisis de asociaciones entre datos fenotípicos y la matriz de datos de marcadores SNP (datos disponibles en Albert y col. 2016) se realizó un estudio exploratorio sobre los contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en los frutos maduros cosechados de las 146 entradas de la población de GWAS. Cada una de las distintas formas de tocoferol presentó un alto grado de variación entre los materiales con coeficientes de variación que van desde 53,5% para α -tocoferol hasta 94,6%

para los contenidos de α -tocotrienol (figura 12 y Tabla S2 suplementaria). Al evaluar los niveles de tococromanoles dentro de cada grupo genético, se puede observar la diferenciación entre estos; por ejemplo para contenidos de α -tocoferol (figura 13), el grupo de SLC muestra las mayores variaciones y un incremento en la acumulación de esta forma con respecto al resto de los grupos llegando en algunos casos a dos veces su concentración. Por otro lado, no se observaron diferencias entre ambientes comparando cada grupo genético globalmente. Sin embargo, para 32 y 43 entradas el efecto ambiental resultó evidente, donde los contenidos de tocoferol total, aumentaron y disminuyeron, respectivamente, asociados a la condición de estrés (figura 12). No obstante, esta observación debe ser tomada con recaudo considerando que el número de individuos analizados para cada entrada en cada condición no permite obtener análisis estadísticos suficientemente robustos para evaluar la significancia de las diferencias.

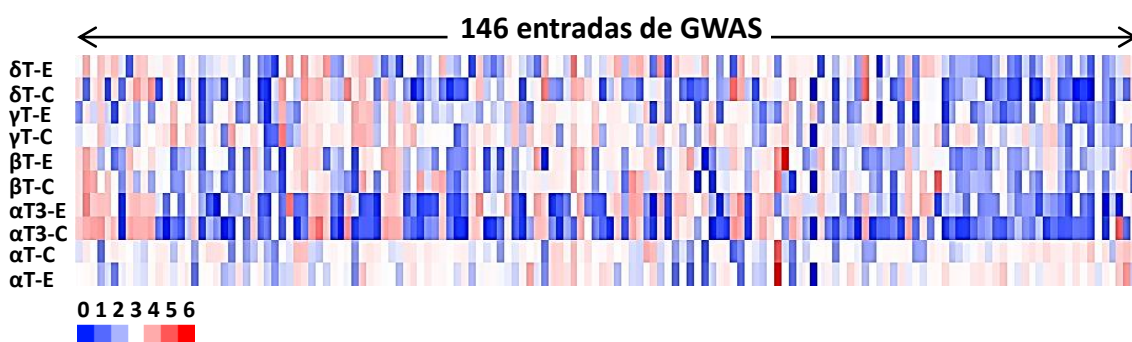


Figura 12. Variabilidad fenotípica para contenidos de tococromanoles en la población de de GWAS estudiada en ambas condiciones ambientales, control (C) y estrés hídrico (E). Para ver una descripción más detallada ver Tabla 2S suplementaria. Isoformas de tocoferoles y tocotrienoles: α T (alfa tocoferol), α T3 (alfa tocotrienol), β T (beta tocoferol), γ T (gama tocoferol). La escala de colores va desde 0 a 6 veces la media de la respectiva forma de tococromanol.

Posteriormente se evaluó la normalidad y homocedasticidad de las distribuciones de los datos fenotípicos, mediante la aplicación de las pruebas de Shapiro Wilks y Levene, respectivamente. Ninguno de los 6 caracteres evaluados presentó normalidad ni homocedasticidad. Se trataron entonces los datos aplicando la transformación de Box Cox (1964).

El análisis de la varianza para los contenidos de tocoferol total y las distintas formas fue homogéneo en cuanto a las proporciones de su distribución (figura 14). Excepción a esta observación fue el caso de α -tocotrienol. Para caracterizar las componentes de esta variabilidad se llevó a cabo un análisis de la varianza (ANOVA), el cual reveló que entre 4 y 11% de la varianza de los contenidos de tococromanoles es explicada por el ambiente, entre 4% y 18% es definida por la interacción genotipo-ambiente (GxE) y el componente

genético explica entre 18,7% a 37% del total, siendo la variable más destacada. Estos análisis también permitieron evidenciar la diferencia entre la distribución de la varianza entre tocoferoles y tocotrienoles, representada a través del α -tocotrienol que presenta un perfil de la varianza completamente distinto, sin componente genético prácticamente. Es importante también remarcar que el lugar de medición tuvo un efecto importante sobre la varianza de los datos (Figura 14).

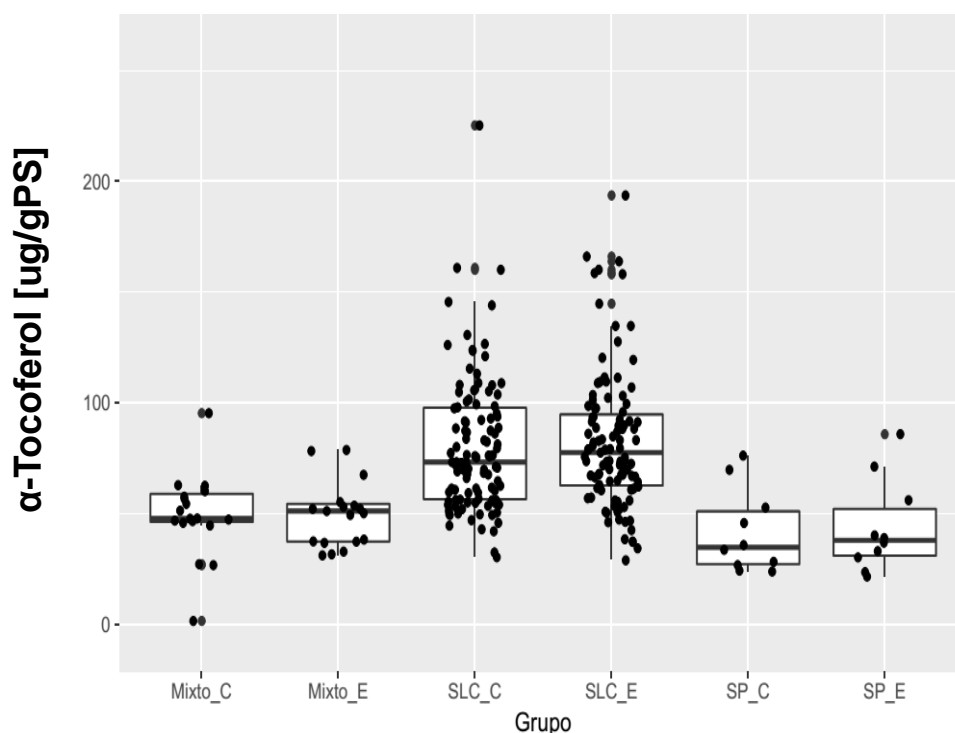


Figura 13. Gráficos de cajas para contenidos de α -tocoferol para los distintos grupos genéticos dentro de la colección de GWAS en condiciones control (C) y de estrés (E). Sobre el eje de abscisas inferior se encuentran referenciados los distintos grupos genéticos: *Solanum pimpinellifolium* (SP); *S. lycopersicum* var. *cerasiforme* (SLC) y un tercer grupo genético fue determinado como “mixto”. Cada una de las cajas describe la distribución de contenidos de α -tocoferol en los tres grupos genéticos y ambos ambientes, para las 146 accesiones evaluadas.

De igual forma que para los materiales de la colección de GWAS, en los materiales de la MAGIC POP también se analizó la distribución y normalidad de los caracteres fenotípicos en frutos maduros de las 124 líneas recombinantes F7 (Tabla S3 suplementaria). La variación fenotípica de estos materiales resultó pronunciada, mostraron amplia diversidad, con coeficientes de variación que van desde 61,1% a 161,70% para β - y δ -tocoferol, respectivamente. Posteriormente se procedió a la transformación de los datos aplicando la función Log_{10} para los 6 caracteres fenotípicos evaluados (figura 15) ya que ninguno de los caracteres presento distribución normal. Finalmente para el mapeo y determinación de QTL, se decidió utilizar ambos set de datos, transformados y crudos, como caracteres fenotípicos independientes.

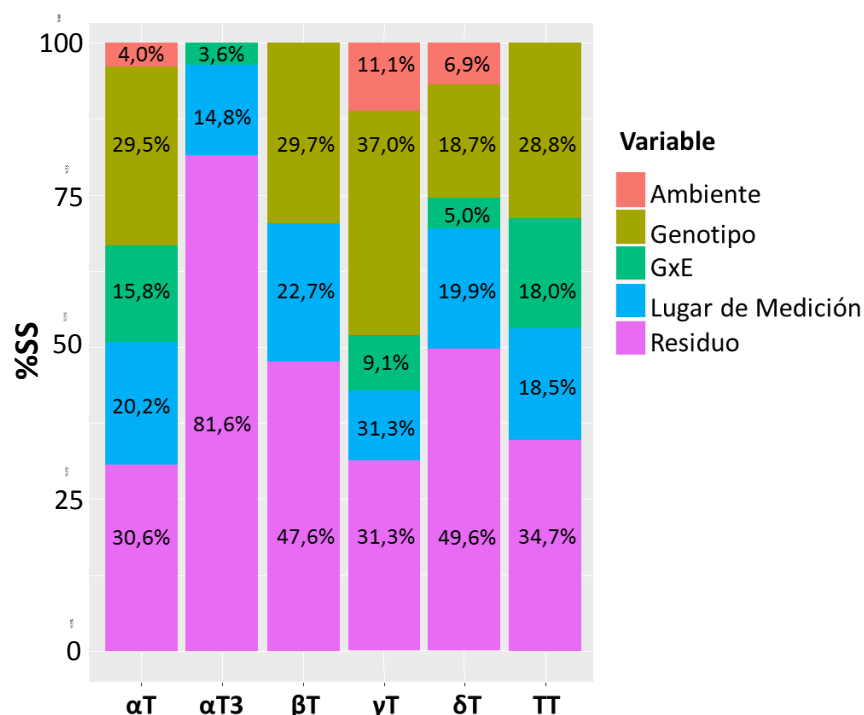


Figura 14. Disección de la varianza fenotípica para los 6 caracteres medidos en el análisis de GWAS. Para cada carácter fenotípico la figura presenta la proporción de cada uno de los efectos en la suma de cuadrados %SS de la varianza total para cada una de las isoformas de tocoferol representadas en naranja la componente ambiente, en ocre el genotipo, verde interacción genotipo-ambiente (GxE), celeste el lugar de medición y fucsia el residuo de la varianza. Isoformas de tocoferoles y tocotrienoles: αT (alfa tocoferol), αT3 (alfa tocotrienol), βT (beta tocoferol), γT (gama tocoferol) y TT (tocoferol total).

II.2.2 Detección de QTL

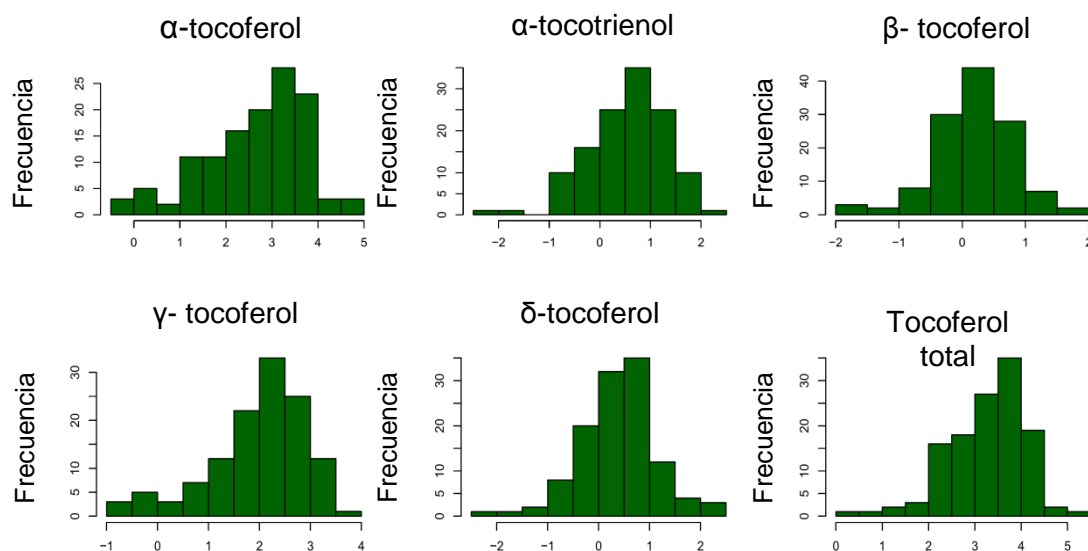


Figura 15. Histogramas de distribuciones fenotípicas para los 6 caracteres medidos en las 124 líneas F7 de la MAGICPOP, posterior a la transformación \log_{10} . Contenidos de α- β- γ- δ-tocoferol, α-tocotrienol y tocoferol total medidos por HPLC expresados como [μg/gPS], ug de la forma de tococromanol por gramo de tejido seco.

II.2.2.1 Por mapeo de asociación

Cuatro análisis de GWAS fueron realizados a partir de los datos fenotípicos de contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en los frutos de los 146 entradas cultivadas en las dos condiciones ambientales. Se seleccionó el MLM que mejor ajustó a cada caso de acuerdo a los test de BIC y extBIC (Schwarz y Gideon 1978). Posteriormente se ajustaron los datos utilizando el paquete de R lme4 (Douglas y col, 2015), específico para la aplicación de modelos lineales mixtos. Se realizaron 4 análisis independientes que difieren principalmente en dos variables: i) el número de pasos o cofactores del modelo que explican la varianza residual asociada al carácter fenotípico; seleccionándose 5 y 10 pasos, lo cual implica, respectivamente una menor y mayor potencia para identificar QTL epistáticos o de efecto pobre y ii) el valor umbral para determinar la significancia de la asociación: seleccionando aquellas que estuvieran por debajo de 10^{-4} y 10^{-5} (ver materiales y métodos para más detalle). De esta manera se realizaron cuatro análisis de GWAS: análisis 1 con 5 pasos un p-valor 10^{-5} ; análisis 2 y 3 con 5 pasos, p-valor 10^{-4} y el análisis 4 con 10 pasos, p-valor 10^{-4} . La diferencia entre los análisis de GWAS 2 y 3 radicó, en que el GWAS 2, los datos no fueron ajustados, solamente normalizados.

Los intervalos físicos que determinaron cada uno de los QTL identificados fueron calculados en base al desequilibrio de ligamiento (LD) entre pares de marcadores tal como se describe en Albert y col. (2017), tomando un valor de ligamiento crítico de 0,28, por encima del cual se considera que dos marcadores están ligados (más detalles se pueden consultar en Materiales y Métodos sección V.1.7). Dependiendo las regiones genómicas donde se identificaron los QTL, los intervalos variaron entre 65,3 Mpb, comprendiendo 2488 genes, hasta unos pocos cientos de pares de bases con asociaciones a un único gen (Figura 16). De esta forma se determinó el número de genes correspondientes a cada intervalo de confianza, para cada uno de los 4 análisis realizados (Tabla 3). Los cuatro análisis de GWAS identificaron un total de 60 asociaciones a SNPs de forma significativa para los contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en frutos, las cuales se encuentran representadas con sus respectivos intervalos de confianza en la Figura 16. Todos los cromosomas mostraron al menos un QTL. En el cromosoma 1 se encontró el mayor número de QTL, con diez asociaciones. Mientras que en el cromosoma 8 solo se encontró una asociación. Un caso particular se dio para el polimorfismo de un SNP que se ubica al comienzo del brazo pequeño del cromosoma 1; una región de 0,3 Mpb donde se encuentran 68 genes. En esta misma región mapean QTL tanto para β - como para γ -tocoferol. Estos resultados son consistentes con el hecho de que la mayoría de los pasos de biosíntesis entre las diferentes formas de tococromanoles son compartidos.

II.2.2.2 Por mapeo de intervalo

Para la detección e identificación de los polimorfismos causales de las variaciones fenotípicas medidas en la población de líneas recombinantes F7 que componen la MAGIC POP se realizó lo que se denomina un mapeo de intervalo simple utilizando la herramienta *mpIM* de R (Huang y George 2011). Esta función computa el efecto de cada parental en base a una regresión de múltiples probabilidades, función *mpprob*, con ventanas de un tamaño de 2 centimorgans, las cuales definen la resolución del mapeo. Para evaluar el grado de significancia del efecto parental para cada uno de los QTL identificados se utilizó el Test de join Wald con un umbral estadístico de p valor $< 1.72 \times 10^{-4}$.

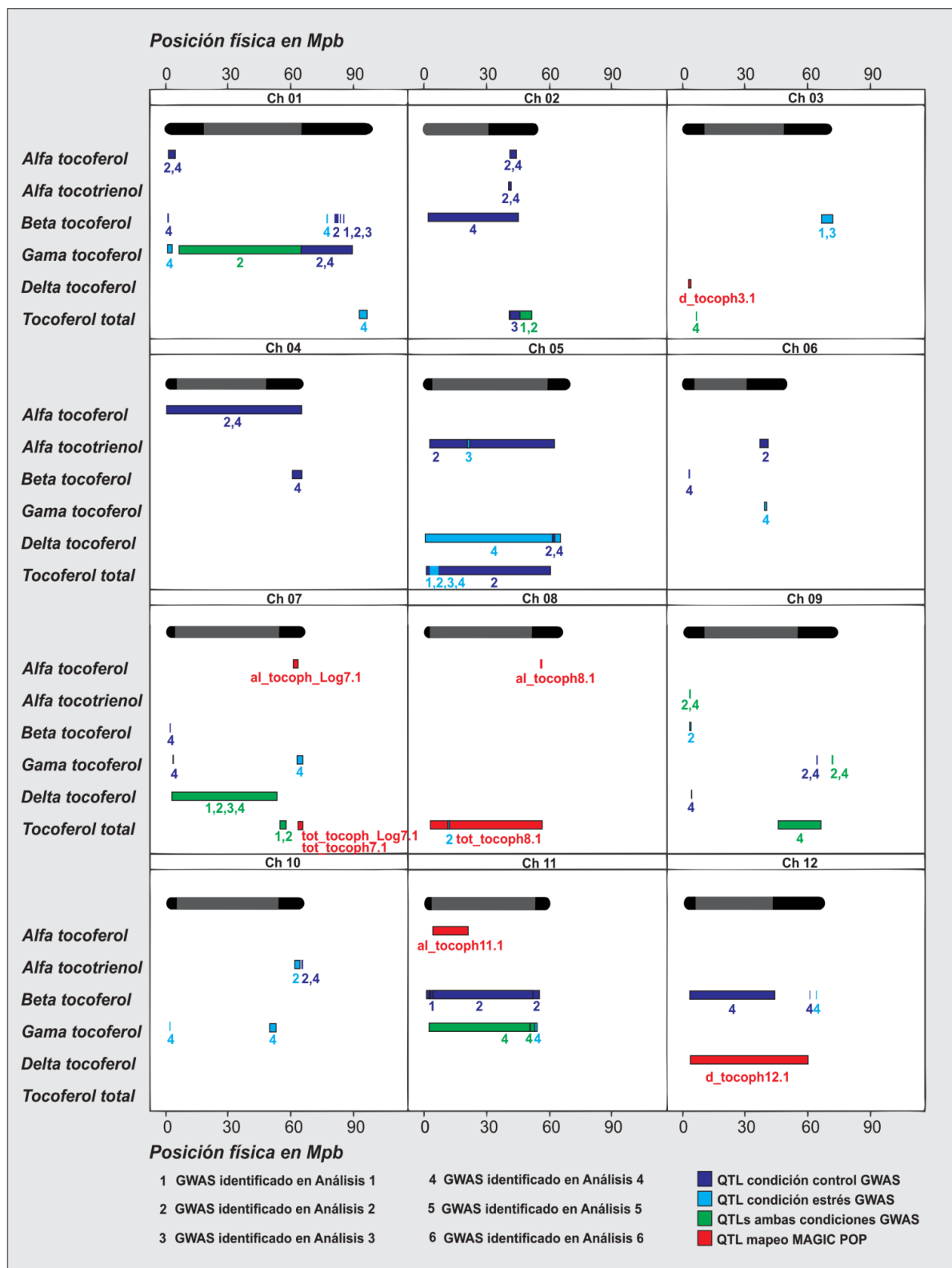


Figura 16. Mapeo físico de los QTL determinados por GWAS y mapeo por intervalo (MAGICPOP). Las distancias están expresadas en millones de pb hechas a escala, sobre 2 ejes de abscisas superior e inferior basadas en el genoma ensamblado versión SL2.5. Para cada cromosoma y carácter, los QTL detectados están indicados con los números 1 a 4 dependiendo el análisis de GWAS, los intervalos físicos de los mismos fueron calculados en base al LD entre marcadores significativos y asociados (para mayor descripción ver sección V). En color azul están los QTL detectados en condición control, en celeste condición de estrés y en verde aquellos QTL detectados en ambas condiciones los cuales se presentan independientes de la misma. En color rojo se encuentran identificados los QTLs identificados a través del mapeo de QTLs de las 124 líneas F7 MAGIC-POP

Este análisis determinó un total de 8 QTL: 3 para α -tocoferol en los cromosomas 7, 8 y 11 con tamaños de 2,4; 0,3 y 17 Mpb respectivamente; 2 para δ -tocoferol en los cromosomas 3 y 12 con tamaños que van de 1,3 en el primer caso hasta 56,8 Mpb y otros 3 QTL asociados a contenidos totales de tocoferoles en frutos, 2 de ellos solapados en el cromosoma 7 con un tamaño de 2,3 Mpb y un tercero en el cromosoma 8 de 17 Mpb el cual se superpone con otro QTL determinado en el análisis de GWAS 2 de tan solo 0,18 Mpb que mapea sobre un único gen, no anotado hasta el momento. Estos QTL se encuentran representados en color rojo en la figura 16. El total de genes candidatos que comprende la sumatoria de los QTL mapeados en los materiales de la MAGICPOP es de 2.018 genes.

Tabla 3. Número de genes en común entre el mapeo de QTL por intervalo obtenido a partir de la MAGICPOP y los 4 análisis de GWAS.

Análisis de GWAS	Total genes IC GWAS^(*)	Numero genes en común al Mapeo de QTL por intervalo
1	3489	0
2	12733	619
3	3616	0
4	14077	1436
^(*) Los distintos intervalos de confianza (IC) determinan la distancia física en los distintos cromosomas asociados a los caracteres medidos y fueron determinados en base al LD entre marcadores ligados. La suma de todos estos intervalos fue tomada para el cálculo total de genes de cada análisis de GWAS.		

Con el fin de evaluar la robustez entre los distintos análisis de GWAS (más y menos astringentes) se compararon el número de asociaciones compartidas entre los análisis. Los estudios presentaron: 6, 27, 13 y 36 asociaciones significativas para los análisis de GWAS 1, 2, 3 y 4, respectivamente. Diecisiete de estas (28,3% del total de asociaciones) fueron compartidas por al menos 2 de ellos. Los análisis 2 y 4 son los que más comparten con un total de 21% en común mientras que los análisis 1, 2 y 3 comparten 10% entre sí. Solo dos del total de asociaciones fueron determinadas en los 4 estudios, estas mapean sobre los cromosomas 5 y 7 para tocoferol total y δ -tocoferol, respectivamente. Otra forma por la cual se analizó la robustez de los análisis fue comparando las listas de genes candidatos determinados en cada análisis, y representado a partir de un diagrama de Venn (Figura 17). En el centro de esta figura se puede apreciar un total de 2.086 genes en común a los cuatro estudios, representando 11,5% del total de genes identificados en común en todos los análisis.

El análisis GWAS 2 presentó la mayor cantidad de genes en común, con un total de 10.499 (57,6%) presentándose como el más reproducible, le siguen en orden los análisis 4, 3 y 1 con 49%, 19% y 18% del total de genes.

A continuación se evaluó cuan efectivo habían sido los 4 análisis de GWAS respecto a la metodología de mapeo por intervalos utilizando la MAGIC POP. Comparando la lista de genes candidatos determinadas por los 4 estudios de GWAS frente a los 2.239 identificados como candidatos mediante mapeo por intervalos (Tabla 3), se observa que el análisis 4 fue el que mayor número de genes comunes presentó seguido del análisis 2. Los otros dos análisis no tuvieron genes compartidos con el mapeo por intervalos. El mayor numero de asociaciones compartidas entre el análisis 2 y 4 y el mapeo de QTL simple podría deberse al mayor número de genes asociados en estos análisis (12733 y 14077, respectivamente) comparado con los análisis 1 y 3 (3489 y 3616, respectivamente).

Otra manera más específica para evaluar la sensibilidad y robustez de estos análisis es evaluar el enriquecimiento de genes involucrados en el metabolismo secundario (MS) y genes relacionados al metabolismo de los tococromanoles

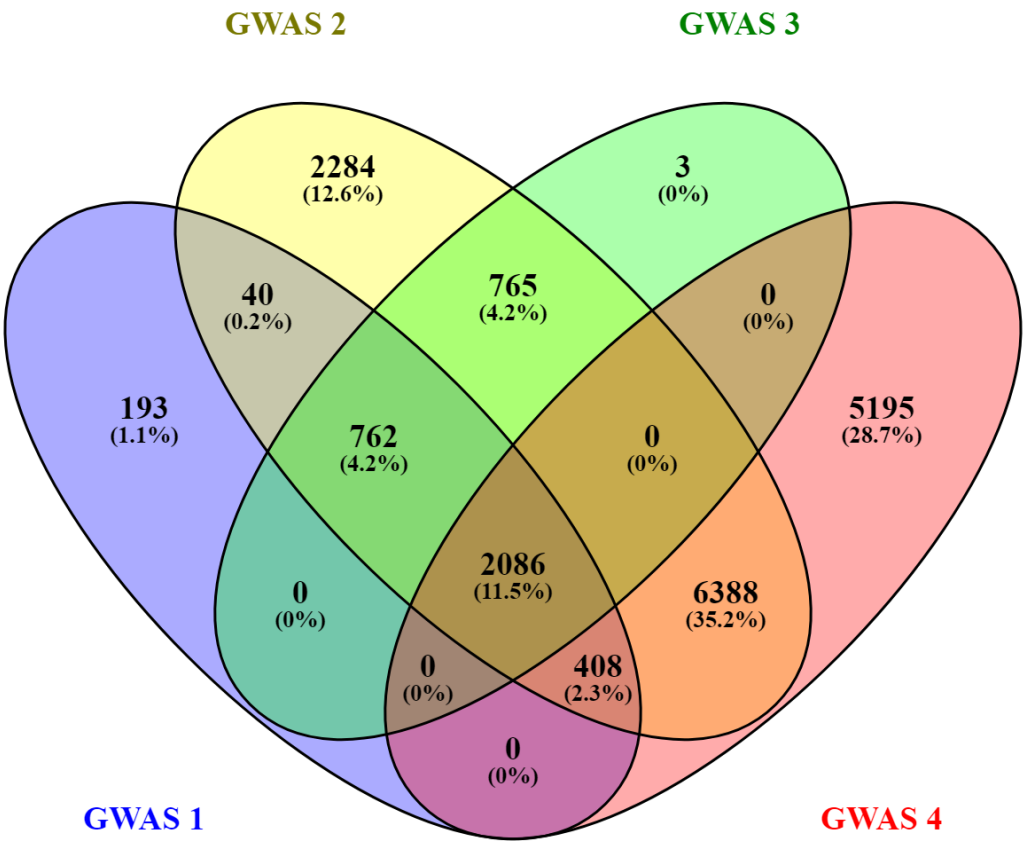


Figura 17. Cantidad de genes comunes a los distintos análisis de GWAS con distintos grados de astringencia mediante un diagrama de Venn. Para las listas de genes candidatos de cada uno de los 4 análisis de GWAS los conjuntos de genes particulares identificados en cada uno fueron representados en colores azul, amarillo, verde y rojo para los análisis 1, 2, 3, 4 respectivamente.

(VTE) (Tablas 4 y 5, respectivamente). Para definir el primer set se tomaron aquellos genes con anotación funcional “Mapman” relacionada a MS. Para el segundo set se elaboró una lista de genes candidatos basado en trabajos publicados por nuestro u otros grupos). Los mismos corresponden a reportes en tomate (Almeida y col. 2011, Guyer y col. 2014). Adicionalmente, otras tres publicaciones utilizadas fueron realizados en maíz (Diepenbrock y col. 2017, Hong Wang y col. 2018, Lipka y col. 2013) y por lo tanto Identificamos los genes ortólogos en tomate mediante el uso del software OrthoDB (Kriventseva y col. 2015) seleccionando aquellos con alta identidad de secuencia (85% de identidad de secuencia). Algunos de ellos han sido validados mediante el análisis de mutantes u otros experimentos de genómica funcional (ver en negrita, Tabla S5).

Los resultados del análisis de enriquecimiento en categorías funcionales no son definitorios ya que las diferencias entre los 4 análisis de GWAS son acotadas. Por otro lado, el mapeo por intervalos a partir de la MAGICPOP presentó un mayor enriquecimiento en genes VTE con 53% por encima de lo esperado para la categoría. Es importante remarcar la estrecha relación inversa que existe entre el enriquecimiento para un análisis dado y el tamaño de los intervalos de los QTL que comprenden las regiones de alto ligamiento génico, los cuales pueden extenderse a varios Mpb en algunos casos. En base a las observaciones comparativas de los 4 análisis de GWAS, evaluando el grado de reproducibilidad y robustez de las asociaciones significativas en común y de los genes candidatos dentro de los IC; en conjunto con el nivel de enriquecimiento de categorías MS y VTE, se decidió explorar en profundidad las 32 asociaciones detectadas por el análisis GWAS 2. Este análisis es el que presentó mayor cantidad de genes relacionados al metabolismo de tococromanos.

Para cada uno de los caracteres evaluados (α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total en frutos maduros) se calculó la heredabilidad para cada una de las condiciones ambientales ensayadas: control y estrés hídrico o ambas en aquellos casos donde la asociación fue constitutiva. También se calculó el porcentaje de la varianza fenotípica explicada por el conjunto de SNPs significativos asociados a ese carácter (PVE) en el paso óptimo del MLMM (Tabla 6). Además se calculó la media de los valores fenotípicos de cada alelo (mayor y menor frecuencia) asociados a cada haplotipo para cada una de las asociaciones identificadas (marcador SNP) para cada condición (estrés, control o ambos ambientes). Estos datos se muestran en la tabla 7, junto al marcador asociado, su grado de significancia; la extensión de cada uno de los IC del QTL y el número de genes comprendidos en cada IC. Por ejemplo para el carácter α -tocoferol se encontraron 3 marcadores asociados en los cromosomas 1, 2 y 4. Los IC detectados a partir de considerar el LD revelaron regiones del

Tabla 4. Enriquecimiento de genes de la categoría Metabolismo Secundario (MS) en cada uno de los análisis aplicados (GWAS y mapeo por intervalos –QTL M-).

Análisis	GWAS 1	GWAS 2	GWAS 3	GWAS 4	QTL M
Total genes por análisis	3.489	12.733	3.616	14.077	2.239
Genes BIN16 (MS).	76	217	72	226	24
Total genes anotados del genoma	36.756				
Total genes BIN16.	524				
Enriquecimiento MS (MS)	1,53	1,2	1,4	1,13	1,33
p-Valor Chi_quadrado	6,3 10 ⁻⁶				
El total de genes anotados y BIN16 MS correspondiente a la categoría MS fue tomado de acuerdo a las librerías de categorías funcionales MapMan (Thimm y col. 2004).					

Tabla 5. Enriquecimiento de genes relacionados al metabolismo de los tococromanoles para cada una de las estrategias de mapeo aplicadas

Análisis	GWAS 1	GWAS 2	GWAS 3	GWAS 4	QTL M
Genes VTE identificados ^(a)	10	36	10	36	8
Total genes Análisis	3.489	12.733	3.616	14.077	2.239
Total genes VTE	86				
Total genes anotados del genoma	36.756				
Enriquecimiento VTE ^(b)	1,21	1,22	1,18	1,09	1,53
p-Valor Chi cuadrado	7,8 10 ⁻¹⁵				
Los genes VTEs considerados para este conforman una lista de 86 genes a partir de 6 trabajos previos de este y otros grupos, tres de ellos en maíz (Diepenbrock y col. 2017, Hong Wang y col. 2018, Lipka y col. 2013) con lo cual previamente se realizó la búsqueda de ortólogos mediante OrthoDB v8, seleccionando solo aquellos que poseían alta identidad de secuencia (Kriventseva y col. 2015) algunos de ellos validados mediante mutantes u otros experimentos, en negrita tabla suplementaria , los otros dos reportes sumados son en tomate (Almeida y Col. 2011,Guyer y col. 2014). El total de genes anotados fue tomado de acuerdo a Categorías funcionales MapMan (Thimm y col. 2004). ^(a) Genes VTE identificados para cada uno de los respectivos análisis. ^(b) Enriquecimiento en genes de la ruta central de VTE.					

genoma de entre 3,2 y 65,3 Mpb, conteniendo un número de genes de entre 329 y 2.488. Si observamos los valores fenotípicos medios para estas 3 asociaciones de α -tocoferol en condiciones control (Figura 18 AyB, Tabla 7), se puede apreciar que no siempre para un alelo dado de mayor o menor frecuencia se observa el mismo efecto de acumulación del isómero en cuestión sugiriendo la presencia de polimorfismos “raros y beneficiosos”, de baja frecuencia y mayor acumulación de α -tocoferol (en comparación con su alelo de mayor frecuencia) como es el caso del polimorfismo asociado al marcador S04_66073220. En contraposición el marcador S02_45438825 presenta un alelo de menor frecuencia (“raro”) y “desfavorable” en cuanto a la acumulación de α -tocoferol.

Tabla 6. Estimación de la heredabilidad y la varianza explicada por cada marcador asociado del análisis GWAS 2.

Carácter	Condición	Heredabilidad ⁽¹⁾	PVE ⁽²⁾	Numero de Asociaciones ⁽³⁾
α -tocoferol	Control	0.455	0.287	3
α -tocotrienol	Control	0.572	0.276	4
α -tocotrienol	Ambos	0.687	0.210	2
β -tocoferol	Estrés	0.376	0.322	2
β -tocoferol	Control	0.613	0.405	4
β -tocoferol	Ambos	0.504	0.353	2
γ -tocoferol	Control	0.485	0.325	4
γ -tocoferol	Ambos	0.469	0.221	1
δ -tocoferol	Control	0.150	0.167	1
δ -tocoferol	Ambos	0.174	0.202	1
Tocoferol total	Estrés	0.239	0.225	2
Tocoferol total	Control	0.450	0.370	3
Tocoferol total	Ambos	0.446	0.294	3

Resumen de asociaciones para contenidos de tocoferoles y α -tocotrienol. ⁽¹⁾La heredabilidad de cada carácter en una condición dada (control, estrés o ambas) para el conjunto de SNPs asociados a la misma fue calculada en el paso 0 del modelo MLM. ⁽²⁾ El porcentaje de varianza fenotípica explicada (PVE) del conjunto de asociaciones para un carácter determinado en una condición dada se estimó en el paso óptimo del modelo y fue expresado entre 0 y 1.

La distribución de los valores medios para cada uno de los 146 genotipos utilizados en el análisis están representados en la figura 18A, la cual presenta la distribución por haplotipo para los tres marcadores asociados (S01_01367023, S02_45438825, S04_66073220) en la condición control. También se puede observar la posición y su grado de significancia en el Manhattan plot de la figura 18B, Los restantes Manhattan plots y graficos de cajas para cada una de las 32 asociaciones identificadas en el analisis de

GWAS 2 se encuentran en la sección anexo figuras S1 y S2, respectivamente. La heredabilidad promedio para los contenidos de las distintas formas de tococromanoles resultó en 0,43 (con un valor mínimo y máximo de 0,15–0,68, respectivamente). Valores similares (H^2 0,30) a estos han sido reportados en tomate (Sauvage y col. 2014) para contenido total de tocoferol. Cabe resaltar que para α -tocotrienol la heredabilidad resultó en valores más altos (0,57 y 0,68 para la condiciones control y “ambas” respectivamente). Algo que resulta remarcable con el modo de herencia de estos caracteres es que esta se reduce marcadamente frente a condiciones de estrés en todos los casos aquí analizados. Por ejemplo para β -Tocoferol y tocoferol total los valores de heredabilidad calculados fueron de 0,613 y 0,45 en condiciones control, en comparación con 0,376 y 0,239, respectivamente en condiciones de estrés.

Por otro lado, los porcentajes de la varianza explicada por el conjunto de SNPs en cada condición rondaron valores entre 16,7% y 40,5% con un promedio de 28,1%. Estos resultados sugieren que los 6 caracteres fenotípicos relacionados al metabolismo de tococromanoles evaluados poseen una arquitectura genética diferencial que se pone en evidencia en la gran diversidad de heredabilidad observada entre tocoferoles y tocotrienoles (estos últimos con mayores valores de heredabilidad). Además al evaluar el porcentaje de la varianza explicada por los SNPs (PVE), en general los valores fueron relativamente bajos, sin embargo para algunos SNPs individuales estos explican mayor proporción. Por ejemplo para γ -tocoferol, una única asociación en ambas condiciones explica 22.1% mientras que en condición control, 4 asociaciones explican tan solo 32,5% de la varianza, mostrando que la arquitectura del carácter es explicada tanto a través de pocos locus de efectos fuertes como así también múltiples variaciones de efecto menor. Resultados similares para otros caracteres metabólicos fueron previamente observados en tomate (Sauvage y col. 2014).

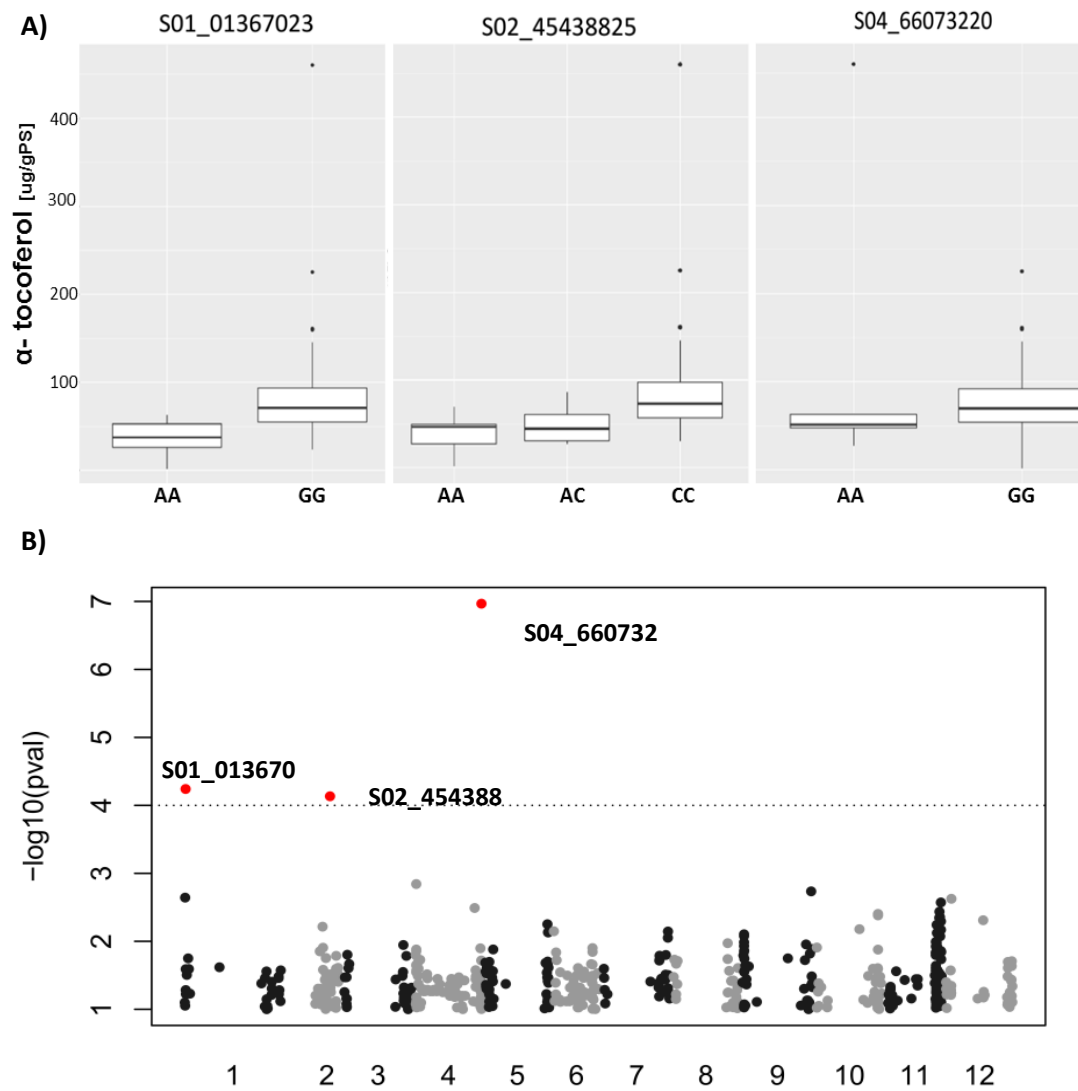


Figura 18. Asociaciones identificadas para α -tocoferol en condiciones control asociadas a cada haplotipo identificadas a partir del análisis de GWAS 2. A) Grafico de cajas para las distribuciones de los 146 genotipos del GWAS según su haplotipo determinado para los tres marcadores asociados S01_01367023; S02_45438825; S04_66073220. **B)** *Manhattan plot* para los tres marcadores significativamente asociados a contenidos de α -tocoferol indicados en color rojo (S01_01367023; S02_45438825; S04_66073220). Se puede observar la línea de corte utilizada en el análisis GWAS 2 con un pValor de 10^{-4} . En el eje de abscisas se puede observar la ubicación de los 6.100 SNPs a lo largo de los 12 cromosomas del genoma.

Tabla 7. Medias fenotípicas asociadas a cada haplotipo por cada una de las asociaciones identificadas en el análisis de GWAS 2 para cada condición; marcador asociado; significancia; intervalos de confianza del QTL; extensión del mismo en Mpb y número de genes

Fenotipo	Condición ⁽¹⁾	Marcador ⁽²⁾	pValor ⁽³⁾	Intervalo de confianza IC ⁽⁴⁾		Alelos	Media Fenotípica ⁽⁶⁾			
							Alelo menor frecuencia		Alelo mayor frecuencia	
				Mpb ⁽⁵⁾	N° genes		C	E	C	E
α-Tocoferol	Control	S02_45438825	4.84E-05	3.2	418	A/C	41	-	83.6	-
α-Tocoferol	Control	S04_66073220	3.62E-08	65.3	2488	A/G	129.9	-	74.5	-
α-Tocoferol	Control	S01_01367023	2.44E-05	3.3	329	A/G	79.6	-	37.4	-
α-Tocotrienol	Control	S06_42560313	1.08E-08	4	514	G/T	5.4	-	6.8	-
α-Tocotrienol	Control	S10_63482478	2.38E-08	0	4	C/G	13.1	-	6.3	-
α-Tocotrienol	Control	S02_48387492	6.15E-06	1.3	178	A/G	12.1	-	4.5	-
α-Tocotrienol	Control	S05_04037126	8.25E-05	0	4	C/T	2.2	-	6.6	-
α-Tocotrienol	Ambos	S10_62693839	4.54E-05	2.5	363	C/T	7.5	7.3	6.3	6.9
α-Tocotrienol	Ambos	S09_00835037	8.53E-05	0.4	58	C/T	13.5	11.9	5.4	6.1
β-Tocoferol	Estrés	S01_85997795	4.18E-07	0.2	17	C/G	-	5.8	-	3.1
β-Tocoferol	Estrés	S03_66764638	1.62E-05	0	1	A/G	4.4	4.3	2.5	2.8
β-Tocoferol	Control	S11_52844288	1.42E-07	54.6	2166	A/G	2	-	3.5	-
β-Tocoferol	Control	S09_01517398	9.58E-06	0.7	74	C/T	1.4	-	3.5	-
β-Tocoferol	Control	S01_85349943	1.40E-05	0	6	A/G	4.3	-	2.5	-
β-Tocoferol	Control	S11_07669410	3.42E-05	48.4	1431	A/C	4.1	-	3.3	-
β-Tocoferol	Ambos	S01_85997795	2.67E-06	0.2	17	C/G	5.1	5.8	3.2	3
β-Tocoferol	Ambos	S03_66764638	1.86E-05	0	1	A/G	4.4	4.3	2.5	2.8
γ-Tocoferol	Control	S05_62738996	9.70E-06	0	4	C/T	24.3	-	16	-
γ-Tocoferol	Control	S09_72090185	2.38E-05	0.4	1	A/G	33.5	-	19.2	-

γ-Tocoferol	Control	S01_96371008	4.26E-05	24.7	1707	A/G	13.4	-	20.3	-
γ-Tocoferol	Control	S09_67892043	8.89E-05	0.2	8	A/T	4.3	-	2.5	-
γ-Tocoferol	Ambos	S01_05364660	6.83E-05	59	815	A/G	27.5	24.3	19.1	15.6
δ-Tocoferol	Control	S05_63053424	6.65E-05	1.2	133	C/G	1.2	-	1.9	-
δ-Tocoferol	ambos	S07_19456301	6.66E-07	50.8	741	C/T	2.6	2.8	1.3	1.5
Tocoferol total	Estrés	S05_05940022	1.25E-06	60.3	1703	A/C	76.1	-	105.4	-
Tocoferol total	Estrés	S08_00602879	8.92E-05	0.8	1	C/T	107.4	-	97	-
Tocoferol total	Control	S02_45438825	2.58E-08	3.2	418	A/C	64.9	-	109	-
Tocoferol total	Control	S04_66073220	7.46E-08	65.3	2488	A/G	165.1	-	99.6	-
Tocoferol total	Control	S05_62738996	2.82E-06	0	4	C/T	118.1	-	88.9	-
Tocoferol total	Ambos	S05_59171718	1.62E-08	58.7	1450	A/G	69.9	63.1	105.2	103
Tocoferol total	Ambos	S07_57109419	3.39E-08	3	231	A/G	142.7	116.1	100.1	98.7
Tocoferol total	Ambos	S02_49774221	5.90E-07	10.4	1396	C/T	72.2	70.7	103	101.6

⁽¹⁾Condición, indica el ambiente en el cual fue asociado cada uno de los 32 marcadores significativos determinados por el análisis de GWAS2.⁽²⁾ Es el identificador asociado a ese SNP dentro de la matriz de 6100 (publicado en Albert y col. 2016). ⁽³⁾Grado de significancia del SNP al carácter fenotípico VTE. ⁽⁴⁾Intervalo de confianza calculado en base al desequilibrio de ligamiento LD, más detalle se encuentra en la sección materiales y métodos V.1.6. ⁽⁵⁾Distancia del intervalo de confianza en megapares de bases ⁽⁶⁾Media fenotípica para el conjunto dentro de los 146 materiales evaluados y para cada uno de los alelos que determinan los distintos haplotipos en las condiciones C (control) y E (estrés).

II. 2.3 Genes identificados por análisis de GWAS y mapeo por intervalo

Las metodologías de mapeo aplicadas en esta tesis permitieron la identificación de 60 QTL a partir de análisis de GWAS (32 de estos explorados en mayor profundidad para el análisis del GWAS2) y otros 8 QTL determinados a partir del mapeo por intervalo en las 124 líneas recombinantes que componen la MAGICPOP. Ambos resultados refuerzan la candidatura de los genes identificados. El listado completo de estos genes, productos en las vías que convergen el flujo de carbono para la síntesis de tococromanoles se puede consultar en la Tabla S5. Un número significativo de estos genes han sido previamente reportados como candidatos para los caracteres estudiados en esta tesis, tanto en tomate (Almeida y col. 2011) como en otras especies (Mach 2015, Diepenbrock y col. 2017; Lipka y col. 2013, Hong Wang y col. 2018). Estos resultados aportan robustez a la detección de QTL y al descubrimiento de nuevos genes candidatos localizados en estos intervalos.

Si bien el catálogo completo de los nuevos candidatos incluye a 32 genes, su análisis detallado exceden los alcances de esta tesis. A modo de ejemplo, se presentan a continuación resultados que permiten posicionar dos de estos genes como posibles blancos para futuros análisis funcionales.

El gen **CM(2)** está codificado en el locus **Solyc11g017240** cuyo producto es la enzima Corismato Mutasa. Esta enzima cataliza el paso de síntesis de corismato a prefenato en la vía del shikimato. Estos intermediarios convergen

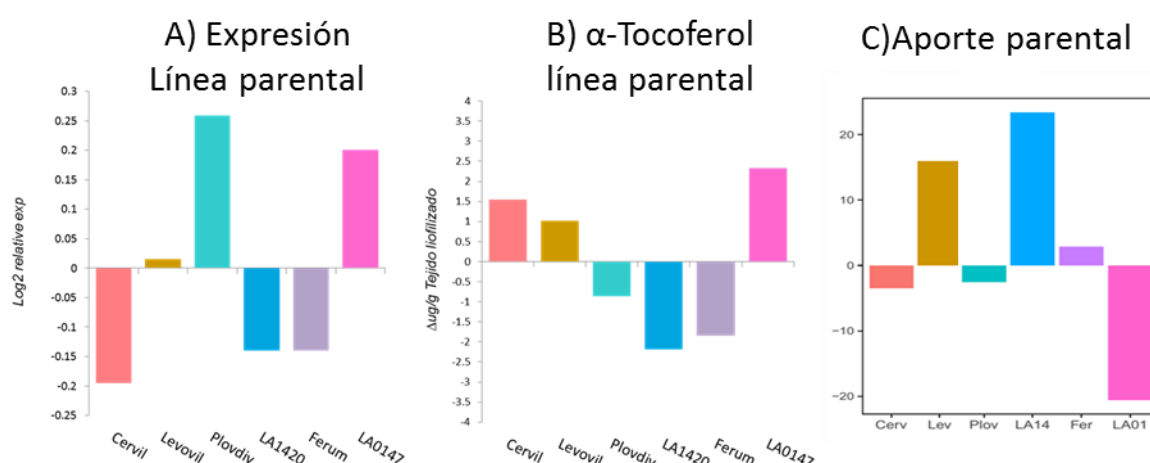


Figura 19. Análisis de CM(2) a partir de datos de expresión, contenidos de tocoferoles y aporte parental de la MAGIC-POP. **A)** Expresión relativa del gen *CM(2)* *Solyc11g017240* que codifica para la corismato mutasa 2 en frutos en estadio de expansión celular en 6 de los 8 parentales de la MAGICPOP: **B)** Contenidos de α -tocoferol en frutos en expansión celular de las líneas parentales **C)** Aporte parental al fenotipo asociado (α -tocoferol) de cada alelo parental sobre la Corismato Mutasa a partir de los datos de mapeo por intervalo. Los parentales están representados en barras de colores de izquierda a derecha: Cervil (coral), Levovil (mostaza), Plovdiv (turquesa), LA1420 (celeste); Ferum (violeta); LA0147 (Fuxia).

para la síntesis de los aminoácidos aromáticos fenilalanina, tirosina y triptófano. A su vez, la deshidrogenación de prefenato aporta el primer intermediario para la síntesis de tococromanoles; el homogentisato. *CM(2)* fue identificado tanto en el análisis de GWAS 2 como a través del mapeo por intervalo sobre un QTL del cromosoma 11 (*al_tocoph11.1*) comprendiendo una región de 11,8 cM (Figura 16). Esta región incluye un total de 619 genes, de los cuales 187 son polimórficos entre algunas de las entradas de GWAS y entre los parentales de la MAGICPOP. Dentro de estos, *CM(2)* es el único gen anotado en la categoría metabolismo secundario de la clasificación de Mapman. Este gen presenta un total de 88 polimorfismos entre los alelos de las 8 líneas parentales MAGICPOP y el genoma de referencia publicado de *S. lycopersicum* (cv.Heinz) (Tomato Genome Consortium, 2012); 28 de estos se encuentran río abajo del gen, 33 río arriba, 25 dentro de intrones, 6 son Indels y 3 SNP sinónimos sobre exones. Uno o más de estos polimorfismos son posibles candidatos a estar asociados a la modulación de la expresión. Explorando perfiles de acumulación de ARN (experimentos de ARN-seq datos aun no publicados Bermudez y col. 2018) de frutos inmaduros en estadio de expansión celular se observa que los niveles de expresión de este gen en los parentales Levovil, LA1420, Ferum y LA0147 (Figura 19A) guardan relación directa con los niveles de α -tocoferol de los frutos en estadio de expansión celular (Figura 19B). A su vez, estos son los parentales que muestran mayor aporte dentro de la población al efecto de este QTL (Figura 19C). Para el caso de los parentales Cervil, Plouvidiv y Ferum el aporte de ambos es mínimo en comparación a los demás parentales, ya que la contribución alélica de este QTL viene dada principalmente a través de los genotipos LA1420, Levovil y LA0147.

Otro de los casos que merece analizarse en detalle es del gen **VTE6** codificado en el locus **Solyc07g062180**. El producto de este gen es la enzima fitil fosfato quinasa que agrega un segundo fosfato al fitil monofosfato, produciendo fitil difosfato que constituirá la cola lipofílica de las moléculas de los distintos tocoferoles (Munné-Bosch y Alegre, 2002). Este gen se localiza dentro de tres de los QTL anidados, (*al_tocoph_Log7.1*) (Figura 16) para contenido de α -tocoferol, *tot_tocoph_Log7.1* y *tot_tocoph71* para contenidos de tocoferol total en los frutos. El primero de estos comprende 292 genes dentro de los cuales 73 se diferencian en su aporte alélico parental, incluido *VTE6* que es el único de la categoría metabolismo secundario. Si analizamos los perfiles de expresión de *VTE6* en los frutos de los genotipos parentales (Figura 20A) se puede observar que existe una correlación inversa (coeficiente de correlación de Pearson $\delta=-0,83$ $p<0,05$) y de igual magnitud con el contenido de α -tocoferol (Figura 20B). Si evaluamos el aporte individual de cada alelo (figura 20C) 7, LA1420 y LA0147 presentan relaciones directas mientras que Ferum y Levovil muestran un aporte en sentido opuesto.

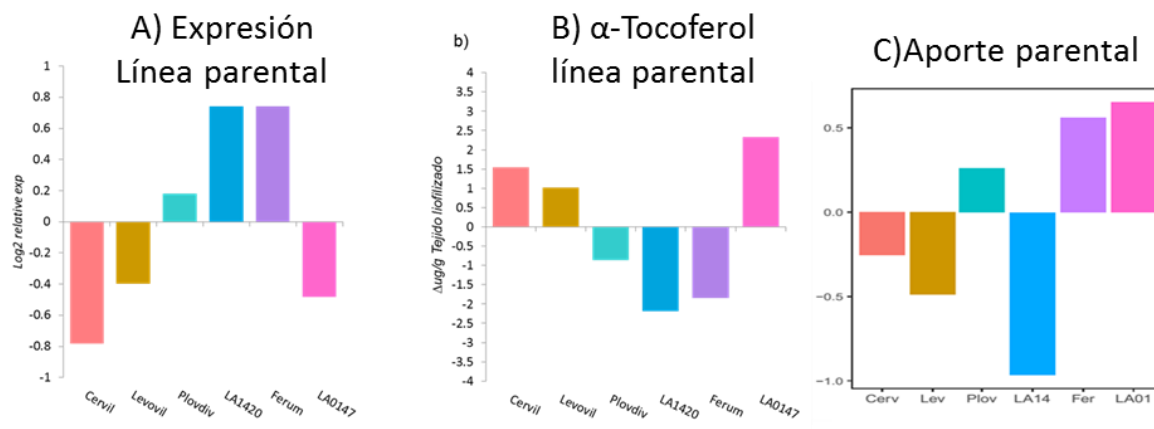


Figura 20. Analisis de VTE6 a partir de datos de expresión, contenidos de tocoferoles y aporte parental de la MAGIC-POP **A)** Expresión relativa de cada uno de los parentales para el gen que codifica para la VTE6, en estadio *cell expansion* celular en 6 de los 8 parentales de la MAGICPOP: **B)** Contenidos de α-tocoferol en frutos en expansión celular de las líneas parentales **C)** Aporte parental al fenotipo asociado (α-tocoferol) de cada alelo parental sobre la VTE6 a partir de los datos de mapeo por intervalo. Los parentales están representados en barras de colores de izquierda a derecha: Cervil (coral), Levovil (mostaza), Plovdiv (turquesa), LA1420 (celeste); Ferum (violeta; LA0147 (Fuscia).

II.2.4 Identificación de factores de transcripción asociados al metabolismo de tococromanoles

Con el objetivo de identificar *loci* codificando factores de transcripción (FT) dentro de los IC asociadas a los caracteres de interés se realizó una búsqueda en dos pasos. Primero se identificaron FT que reconozcan sitios regulatorios en *cis* en el grupo de 86 genes relacionados al metabolismo de tococromanoles mencionados más arriba y se evaluó cuáles de estas interacciones estaban sobrerrepresentadas. En una segunda etapa se analizó cuál de estos FT se encuentra dentro del grupo de candidatos identificados por el análisis GWAS 2 y el mapeo por intervalo con la MAGICPOP. Utilizando el algoritmo contenido en la base de datos PlantTFDB (Jin y col. 2017) se identificaron aquellos FT cuya interacción con genes VTE resultaron sobre-representados en el conjunto de genes encontrados en los IC. Para verificar el grado de significancia de este enriquecimiento, se utilizó la prueba exacta de Fisher, contrastando contra todos los motivos de la base de datos de regulaciones previamente reportadas (para más detalle, ver materiales y métodos sección V.1.9). A partir de este análisis se identificaron 767 posibles regulaciones entre 211 FT sobre 84 de los 86 genes VTE (Tabla S5 suplementaria). De estos 211 FT, 25 de ellos podrían reconocer sitios regulatorios en *cis* en las regiones promotoras de 84 de los genes VTE (Tabla 8).

Tabla 8. Factores de transcripción sobrerrepresentados en la vía de tococromanos.











ID de TFs ^(a)	Genes totales ^(b)	Genes c sitios cis ^(c)	Genes ^(d)	gVTE c/unión cis ^(e)	p_valor ^(f)	FDR ^(g)
Solyc10g055410	34727	995	84	9	1.4E-04	3.0E-02
Solyc11g011050	34727	898	84	8	3.2E-04	3.3E-02
Solyc05g014290	34727	949	84	8	4.8E-04	3.3E-02
Solyc10g005760	34727	670	84	6	1.2E-03	6.3E-02
Solyc09g008390	34727	350	84	4	1.6E-03	6.8E-02
Solyc03g098320	34727	537	84	5	1.9E-03	6.8E-02
Solyc09g090130	34727	2795	84	14	2.7E-03	8.1E-02
Solyc01g094360	34727	1317	84	8	4.5E-03	1.2E-01
Solyc10g079050	34727	747	84	5	9.5E-03	2.1E-01
Solyc01g065540	34727	188	84	2	1.1E-02	2.1E-01
Solyc02g036370	34727	552	84	4	1.1E-02	2.1E-01
Solyc12g006800	34727	201	84	2	1.2E-02	2.2E-01
Solyc10g084370	34727	584	84	4	1.3E-02	2.2E-01
Solyc01g109880	34727	868	84	5	1.8E-02	2.7E-01
Solyc10g005550	34727	1398	84	7	1.9E-02	2.7E-01
Solyc11g065840	34727	2303	84	10	2.2E-02	3.0E-01
Solyc10g044680	34727	1809	84	8	3.1E-02	3.4E-01
Solyc05g015850	34727	489	84	3	3.1E-02	3.4E-01
Solyc04g074170	34727	1814	84	8	3.1E-02	3.4E-01
Solyc09g066010	34727	536	84	3	4.1E-02	3.9E-01
Solyc08g081610	34727	545	84	3	4.3E-02	3.9E-01
Solyc06g005680	34727	326	84	2	4.4E-02	3.9E-01
Solyc01g090840	34727	334	84	2	4.7E-02	3.9E-01
Solyc01g108080	34727	336	84	2	4.8E-02	3.9E-01
Solyc09g014990	34727	570	84	3	4.9E-02	3.9E-01
^(a) SolycIDs de todos los factores de transcripción con motivo de unión a las regiones regulatorias de los genes relacionados con la vía de los tococromanos (tabla suplementaria S5). ^(b) Total de genes anotados en el genoma del tomate para la inferencia del análisis de enriquecimiento. ^(c) Total de FT en la base de datos PlantTFDB que poseen sitios de unión al listado de 86 genes VTE. ^(d) Cantidad de genes VTE (input). ^(e) Total de genes con motivos de unión al FT ^(a) , dentro del listado 86 genes. ^(f) p-valor Test exacto de Fisher. ^(g) q-valor o p-valor ajustado a los falsos positivos FDR del Test exacto de Fisher.						

En una segunda etapa se evaluó cuáles de estos FT se encuentran en los IC de los análisis de GWAS 2 del mapeo por intervalo. Un total de 87 FT se encontraron dentro del grupo de 211 que son capaces de reconocer sitios en *cis* de los promotores de los genes VTE. De estos 87, un total de 10 FT se encontraron dentro de los 25 FT con interacciones sobrerrepresentadas

estadísticamente (Tabla 9). Paralelamente, del listado de 2.239 genes identificados en el mapeo por intervalos, cinco se encuentran dentro del grupo de aquellos 211 FT capaces de reconocer sitios regulatorios en los genes VTE. Sin embargo, ninguna de estas posibles interacciones resultó significativa cuando se ensayó su sobrerrepresentación.

Entre los 10 FT identificados en el análisis de GWAS 2, siete de ellos pertenecen a la familia de los MYB y poseen motivos de unión a los promotores que codifican para los siguientes genes VTE: *DHQS* (3-dehydroquinato sintasa), *SDH* (shikimato deshidrogenasa), *PDS* (fitoeno desaturasa), *GGPS*(3) (geranylgeranyl pirofosfato sintasa), *PLSCR* (Scramblase de fosfolípidos), y *EPSPS* (5-enolpiruvilshikimato-3-P sintasa). Otro de estos FT pertenece a la familia de ARR-B (que contienen un dominio del unión al ADN del tipo MYB) y reconocería un sitio de unión al gen *SINE*(2). Un FT de la familia HD-zip C2H2 reconocería un sitio regulatorio en el promotor del gen de la *CM*(1) (corismato mutasa 1) y otro de la familia WRKY con unión a la región promotora del gen *DXS*(2) (1-Deoxy-D-xilulosa-5-P sintasa). El hecho que la mayoría de los FT identificados pertenezcan a la familia de los MYB, reconozcan elementos en *cis* ya descritos en genes de plantas de tomate (Wang y col. 2005), que estas interacciones posean alta preferencia posicional y que estén sobre-representados en los promotores de genes VTE, apoya la hipótesis de que estos elementos son reguladores funcionales (Tabla 9).

Tabla 9. Factores de transcripción y motivos regulatorios en *cis* sobrerrepresentados en la vía de los tococromanos e identificados en el análisis de GWAS 2.

ID ⁽¹⁾	Familia del TF ⁽²⁾	Gen VTE target ⁽³⁾	Motivo <i>cis</i> ⁽⁴⁾	e-valor ⁽⁵⁾	Comienzo del sitio ⁽⁶⁾	Termina el sitio ⁽⁷⁾
Solyc11g011050 thm16	MYB R2R3MYB43	DHQS Solyc02g083590		1.50E-15	14	61
				2.80E-16	71	111
Solyc11g065840 SIMYB46	MYB R2R3MYB	SDH Solyc02g088460		8.40E-15	32	79
				6.00E-16	85	129
Solyc05g014290	MYB R2R3MYB	PDS Solyc02g080620		2.80E-16	71	111
				3.50E-13	14	61
Solyc09g008390 SIMYB22	MYB R2R3MYB	GGPS(3) Solyc02g085700		1.10E-15	17	64
				3.00E-17	73	115
Solyc01g094360	MYB	DHQS Solyc02g083590		6.10E-18	14	61
				3.90E-15	67	112
Solyc01g065540	ARR-B	SINE2 Solyc03g062700		6.10E-29	211	264
				1.60E-27	31	139
Solyc10g084370	MYB	Scramblase Solyc03g006460		3.20E-14	35	79
Solyc05g015850 SIWRKY75	WRKY	DXS(2) Solyc01g028900		2.00E-29	83	150
				1.50E-33	93	151
Solyc04g074170	MYB	EPSPS Solyc01g091190		3.60E-19	14	61
				1.40E-16	67	112
Solyc01g090840	C2H2 HDzip	CM(1) Solyc02g088460		1.60E-03	7	27
				6.50E-04	182	204
				6.40E-04	233	255

⁽¹⁾Identificador SolyclD del factor de transcripción y sitios de unión a genes VTE sobrerrepresentados reportado en tomate en (Wang y col. 2005)

⁽²⁾Familia del factor de transcripción FT identificado. ⁽³⁾Gen blanco reportado en 6 trabajos distintos, tabla S5. ⁽⁴⁾Secuencia nucleotídica del motivo de unión CIS para cada uno de los TF a los genes VTE blanco, ⁽⁵⁾Significancia de la sobrerrepresentación de la interacción ⁽⁶⁾Posición de inicio del motivo con respecto al inicio de la traducción ⁽⁷⁾ Posición de finalización del motivo con respecto al inicio de la traducción.

II. 3. Discusión

Los QTL identificados en este capítulo permitió generar un conjunto crítico de genes candidatos potencialmente involucrados en la biosíntesis y regulación de tococromanoles en frutos maduros de tomate. Más aun, dado que la variación fenotípica fue medida tanto en condiciones normales (riego) como de estrés hídrico, el impacto el ambiente en la heredabilidad y arquitectura genética de estos caracteres pudieron ser evaluados. A su vez, el mapeo de QTL a partir de una población multiparental MAGICPOP constituyó una estrategia complementaria para la detección de QTL. A partir de los análisis genéticos implementados fue posible reconstruir las vías metabólicas cuyo flujo de carbono confluye en la síntesis de tococromanoles en los frutos de tomate. Esto se realizó a partir del análisis detallado de los 60 QTL identificados de los cuales 32 fueron explorados con mayor detalle. Al mismo tiempo, la identificación de FT posiblemente involucrados en mecanismos regulatorios de la vía constituye una novedad para el estudio de estos caracteres en esta especie. Además del análisis detallado de los dos candidatos presentados más arriba, otros 42 genes reportados fueron identificados por las metodologías de mapeo. En la figura 21, se puede observar la posición de los productos de estos genes en el contexto de las vías biosintéticas de tococromanoles. Algunos de estos genes ameritan ciertas consideraciones. Por ejemplo, el caso de los genes *DXS* (*1y2*), cuyo producto se encuentra al comienzo de la vía del metil eritritol fosfato, ambos parálogos de tomate fueron determinados en estos análisis y *DXS*(2) presentó motivos de reconocimiento por un FT de la familia WRKY (*SIWRKY75*) en su región regulatoria. Este FT es reconocido por su respuesta al estrés (Wang y col. 2005). El ortólogo *AtDXS* de *Arabidopsis thaliana* fue expresado ectopicamente en arroz, resultando en lo que se conoce como *el Golden Rice 2* por su color y acumulación de carotenoides (Bai y col. 2016), demostrando un rol crítico para el abastecimiento de precursores de isoprenoides, particularmente de geranil-geranil-difosfato.

La totalidad de los estudios reportado sobre este gen fueron enfocados en el aumento de carotenoides mediante su sobreexpresión. Por ejemplo, en tomate se produjo un aumento de 1,6 veces en el contenido de carotenoides (Enfissi y col. 2005). Estos mismos resultados se reprodujeron en papa con incrementos en carotenoides de hasta dos veces frente a los controles (Morris y col. 2006). También en sorgo, mediante una triple transgénica expresando ectópicamente *AtDXS*, *CRTI* (*desaturasa de caroteno*) y *PSY* (fitoeno sintasa) se observaron incrementos en todos los trans- β -carotenos desde 2,5 a 9,1 $\mu\text{g.g}^{-1}$ (Che y col. 2016). Por otro lado, para la vía del SK que produce homogentisato como sustrato limitante para la síntesis de tococromanoles, fueron mapeados un gran número de los genes conocidos reportados en la literatura, entre estos: *DAHPS* (3-Deoxi-D-arabino-heptulozanto-7fosfato sintasa); *DHQS* (3-Dehidroquintato

sintasa), *SDH* (shikimato deshidrogenasa); *SK* (shilimato quinasa); *EPSPS* (5-enolpiruvilshikimato-3-fosfato sintasa); *CM(1y2)* (corismato mutasas 1 y 2). Otro gen de otra vía confluyente el cual ha sido reportado recientemente en maíz (Diepenbrock y col. 2017) es la prefenato deshidratasa ADT/PDH, la cual está relacionada a la síntesis de los aminoácidos fenilalanina, la tirosina y el triptófano a partir de la producción de fenilpiruvato.

Dentro de lo que es la ruta central de la biosíntesis de tococromanoles se encontraron los genes parálogos *HPPD(1y2)* para los cuales también hay reportes de plantas transgénicas de arroz con niveles alterados de tocotrienoles, lo cual sugiere que el cuello de botella en la regulación de tocoferoles se encuentra río abajo de la vía. Por otra parte, estas plantas presentaron alterada la relación α -/ γ tocoferol (Farre y col. 2012). También para la vía central de VTE, se encontró asociado el gen que codifican para *VTE2*, enzima encargada de la condensación de homogentisato y fitildifosfato. Para esta enzima, el ortólogo de manzana (*Malus domestica*) fue utilizado para la obtención de plantas transgénicas de tomate, las cuales presentaron incrementos de 3,6 y 2,9 veces para α - y γ -tocoferol en hojas y en frutos 1,7 y 3,7 para las mismas isoformas (Seo y col. 2011). Por otro lado plantas transgénicas de *VTE2* de *Latuca Sativa L* y *A. thaliana* aumentaron dos veces los contenidos de tocoferol en hojas debido principalmente al constituyente γ -tocoferol, mientras que los niveles de clorofila decrecieron hasta 20% mostrando una competencia por sustratos intermediarios. Sin embargo, en el mismo trabajo mutantes para *VTE1* incrementaron 35% los niveles de clorofila. Esto demuestra cómo se altera los contenidos de clorofila según el gen *VTE* introducido (Lee y col. 2007). Por otro lado, un número de trabajos que relacionan la variación a los contenidos de tocoferol estrechamente ligado a la degradación de clorofilas ha crecido de forma exponencial tanto en el estudio de plantas transgénicas como a través de la identificación de los mismos por mapeos de QTLs como GWAS (Almeida y col. 2015, Hong Wang y col. 2018, Mach 2015, Diepenbrock y col. 2017). Tal es el caso de *VTE5*, el cual no fue identificado en estos análisis pero es conocido su importante rol en la producción de fitildifosfato a partir del reciclado de clorofila. Más del 80% del contenido de VTE depende de los niveles de expresión de *VTE5* en frutos y hojas de tomate (Almeida y col. 2015). Presentándose con estos resultados la mayor importancia del reciclado de intermediarios metabólicos que su síntesis de *nov*o. En este trabajo demostraron como su deficiencia impacta en la acumulación de almidón y baja asimilación de CO₂ en hojas. Exponiendo una interacción entre el metabolismo de carbono y lípidos, no reportada hasta el momento. No obstante, nuestro estudio identificó dos genes que codifican para las enzimas del paso anterior y posterior del reciclado de PDP. La enzima quinasa *VTE6*, que aporta el segundo fosfato en el reciclado de PDP, ha sido originalmente reportada en mutantes de *Arabidopsis* a las cuales se le introdujo

el gen VTE6 de *E. coli*, reconstituyendo su funcionalidad e incrementar el fitol difosfato y los contenidos de tocoferoles (Vom Dorp y col. 2015). También PPH (feofitinasasa feoforbide hidrolasa), enzima del paso anterior a VTE5, es la encargada de hidrolizar la clorofila para proveer fitol para la síntesis de PDP (Chander y col. 2008, Zhang y col 2013). Plantas transgénicas de esta enzima mostraron incrementos en contenidos de tocoferoles de semillas de sobreexpresantes. Ambas identificaciones refuerzan las conclusiones de Almeida y col. (2015).

II.4 Conclusiones

Los resultados en conjunto de este capítulo permitieron ver que el orden entre los tres grupos genéticos evaluados dentro de las 146 entradas estudiadas (SL, SP y mixto) del panel de GWAS fue el mismo que reportado por Sauvage y col. (2014) a partir de 163 accesiones que incluyen estos materiales aquí estudiados pero para contenidos totales, ambos resultados son comparables ya que α -tocoferol constituye entre un 80% y 90% del contenido total de tocoferol.

Además identificamos nuevos determinantes asociados a los amplios rangos de variación de contenidos de tococromanoles en frutos maduros de tomate de las poblaciones de GWAS y MAGIC-POP. Se identificaron 60 asociaciones significativas entre SNPs para los 6 caracteres fenotípicos evaluados dentro de una población de 146 genotipos evaluados bajo dos condiciones a través de 4 análisis de GWAS con distinto grado de astringencia utilizando modelos MLM. Posteriormente las comparaciones de las asociaciones en común entre análisis y en el nivel de enriquecimiento para categorías de genes determinó que la puesta a punto para el estudio de GWAS 2 se presentó como el más enriquecido y representativo del conjunto de análisis. Además esto demostró que partiendo de un mismo set de datos fenotípicos y genotípicos, cambios en la astringencia del modelo MLM del GWAS tiene un gran impacto sobre el resultado, extrapolando al número de asociaciones dado que ciertas partes del genoma poseen largos bloques en LD, e incrementan en el número de dichas asociaciones sustancialmente. Paralelamente el análisis de mapeo de QTL de 124 líneas de tomate F7 MAGIC-POP determinó 8 QTLs para α - δ -tocoferol y tocoferol total. La comparación de resultados de ambas metodologías fue complementaria, hubo cierto grado de solapamiento entre los QTL que determinan los contenidos de tocoferol en fruto de estos materiales, un 28% de los genes del mapeo de QTL son compartidos con el listado de genes del GWAS 2. La comparación de los genes candidatos identificados a través de ambas metodologías permitió identificar 32 genes previamente reportados relacionados al metabolismo de tococromanoles lo cual refuerza su candidatura y alienta la caracterización de otros genes candidatos comprendidos dentro de los QTLs aquí reportados. Futuros ensayos podrían enfocarse en la

caracterización de alguno de los 10 FT sobrerrepresentados para genes VTE o genes poco estudiados como mucho de los candidatos aquí identificados que son ortólogos de genes en maíz también asociados a los niveles de tocromanoles en esa especie (Hong Wang y col. 2018).

Es importante resaltar que no obstante la identificación de nuevos candidatos reportados en este capítulo, los caracteres aquí estudiados presentan una alta influencia ambiental (Figura 14) y una moderada o baja heredabilidad como ya había sido reportada previamente para tomate en otros materiales (Sauvage y col. 2014). Adicionalmente, los valores de heredabilidad varían dramáticamente en relación con el ambiente, pasando de 0.613 y 0.45 para β -tocoferol y tocoferol total respectivamente en condiciones de cultivo control a 0.376 y 0.239, en estrés hídrico. La fuerte dependencia del ambiente en la heredabilidad de estos caracteres representa un desafío para incrementar los contenidos de estos compuestos nutraceuticos en tomate. Por lo tanto, una mejor comprensión de las bases genéticas que operan en la determinación de los contenidos de tococromanoles es necesaria, lo cual constituye la motivación del próximo capítulo de esta Tesis.

CAPITULO III- Estudio de los mecanismos de interacción génica y alélica en la regulación del metabolismo de tococromanoles

III.1 Introducción

Los resultados del análisis genético presentado en el capítulo anterior muestran que la heredabilidad promedio de los contenidos de las distintas formas de tococromanoles es de 0,43 (con un valor mínimo y máximo de 0,15 - 0,68; respectivamente), valores inferiores inclusive han sido reportados para tomate (H^2 : 0,30; Sauvage y col. 2014). Más aun, la heredabilidad para los contenidos totales de tocoferol total se reduce drásticamente en plantas cultivadas en condiciones estresantes (0,45 en condiciones de cultivo control comparado con 0,23 en condiciones estresantes). Esta relativamente baja heredabilidad podría deberse a una gran sensibilidad de este carácter al ambiente. Estudios en olivo durante tres años consecutivos demuestran como las condiciones ambientales de déficit hídrico e incremento de la temperatura, afectan los niveles de expresión de varios genes estructurales de la vía de biosíntesis de tococromanoles (Fitol quinasa: *VTE5*, Geranilgeranil reductasa: *GGDR*, 4-Hidroxifenilpiruvato dioxigenasa: *HPPD*, Homogentisato fitil transferasa: *VTE2*, Homogentisato geranilgeranil transferasa: *HGGT*, Dimetil-fitolquinol metil transferasa: *VTE3*, Tocoferol ciclasa: *VTE1* y γ -Tocoferol C-metil transferasa: *VTE4*) (Georgiadou y col. 2015, 2016). Otros de los componentes de las variaciones fenotípicas observadas en el carácter podrían estar dados por efectos de interacción (epistasia) entre distintos *loci* involucrados en la regulación de la vía. Estos efectos, no entendidos aún, podrían también explicar la reducida heredabilidad del carácter.

Estudios previos del grupo permitieron identificar 16 QTL para variaciones en los contenidos de tocoferoles, los cuales co-segregan con la presencia de alelos silvestres de genes codificantes para enzimas de la biosíntesis de tococromanoles (Almeida y col. 2011). En particular, el estudio de una línea de introgresión (IL9.2.6), conteniendo un segmento genómico de la especie silvestre *Solanum pennellii* en el cromosoma 9 de la especie cultivada *Solanum lycopersicum* (cultivar M82), mostró un incremento del 70% en los contenidos de α -tocoferol y tocoferol total en frutos maduros de tomate. La reducción de la introgresión (mediante un retrócrucía con el padre donante y posterior autofecundación) a un segmento de aproximadamente 769 Kpb, comprendiendo 36 genes cuya expresión es detectable en algún tejido de la planta de tomate, permitió la identificación del gen que codifica para la enzima 2-metil 6-fitolquinol metiltransferasa (*VTE3*). Esta enzima transfiere grupos metilo en la reacciones de conversión de 2-metil-6-fitol-1,4 hidroquinona

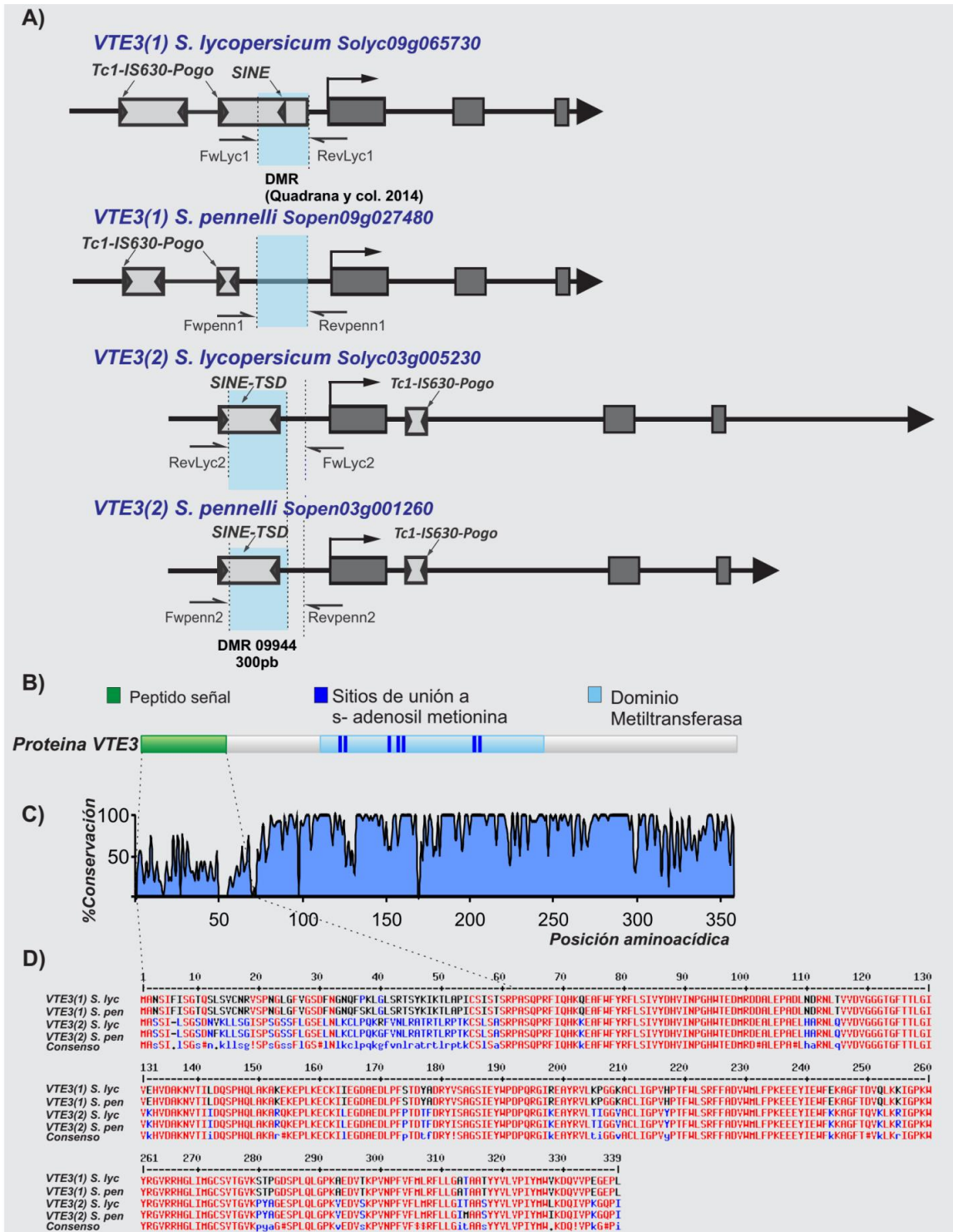


Figura 22. Análisis de secuencias aminoácidas y nucleotídicas de los genes parálogos VTE3(1y2) en las especies *S. lycopersicum* y *S. pennelli*. **A)** Modelo en escala de la estructura génica de los genes parálogos VTE3(1y2) para *S. lycopersicum* y *S. pennelli*. Las cajas grises oscuras representan los exones y las líneas negras que los conecta intrones, las cajas grises claras TEs identificados *in silico*, Tc1-IS630-Pogo, SINE y SINE-TSD. Las flechas flanqueantes a la DMR reportada en Quadrana y col. 2014, representan los primers específicos de alelos utilizados para el análisis de grado de metilación de la región promotora con la enzima McrBc (véase materiales sección V.5), de igual forma para la DMR 09944 reportada en <http://solgenomics.net>. **B)** VTE3(1y2) a nivel proteico, péptido señal de direccionamiento a cloroplasto (verde), dominio metiltransferasa (celeste), sitios de unión a s-adenosilmetionina. **C)** Grado de conservación a nivel AA entre VTE3(1y2) *S. Lycopersicum* para los 339aa. **D)** Multialineamiento de la secuencia de AA entre VTE(1y2) *S. pennelli* y *S. lycopersicum*, secuencia idéntica (rojo), secuencia con polimorfismos en común para VTE3(1) (negro), secuencia con polimorfismos en común para VTE3(2) (azul).

(MPBQ) a 2,3 dimetil-6-fetil-1,4- hidroquinona (DMPQ) que es sustrato para la siguiente enzima, tocoferol ciclasa (*VTE1*) y de 2-metil-6-solanil-1,4 benzoquinona (MSBQ) a plastoquinona (Quadrana y col. 2014).

La proteína codificada en el *locus VTE3(1)* consta de 339 aminoácidos y el gen posee una longitud total de 2.803 pb y un transcripto maduro de 1.317 nt (Figura 22A). Estudios *in silico* y funcionales demostraron que además de producir una proteína de localización en cloroplastos (Figura 23A), cambios en los niveles de expresión de *VTE3(1)* afectan cuantitativa y cualitativamente la producción de tocoferol. Análisis detallados de los mecanismos de regulación transcripcional de *VTE3(1)* demostraron que este gen es regulado epigenéticamente mediante la metilación diferencial del ADN de un retrotransposon no-LTR de la familia SINE localizado en el promotor de este gen (Quadrana y col. 2014 y Figura 22A). De esta manera, bajos niveles de metilación del promotor de *VTE3(1)* se asocian con altos niveles de expresión del gen. Llamativamente, los niveles de metilación de la secuencia promotora de este gen en plantas de tomate cultivadas a campo son significativamente menores que plantas del mismo genotipo cultivadas en condiciones controladas. Estas disminuciones en los niveles de metilación del promotor de *VTE3(1)* son acompañados por aumentos en la expresión del gen, indicando que la respuesta transcripcional del gen *VTE3(1)* tiene una componente epigenética (Quadrana y col. 2014).

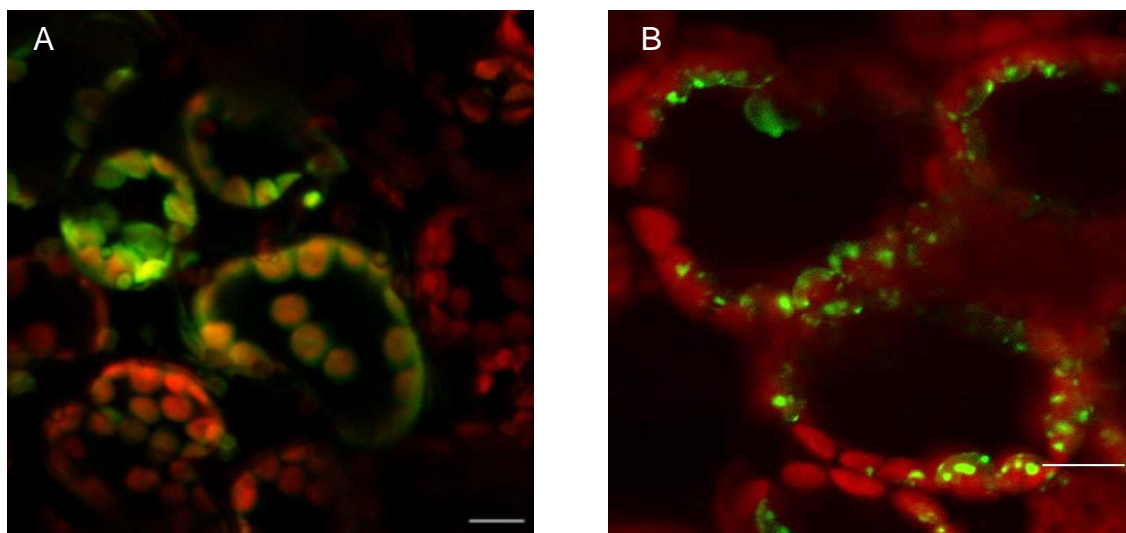


Figura 23. Localización sub-celular de las proteínas VTE3(1) (A) y VTE3(2) (B) de *S. lycopersicum* en hojas de *N. benthamiana*. Se clonó el cDNA completo para cada uno de los genes paralogos y se fusionó con GFP en su extremo carboxilo terminal en el vector pK7FWG2. Se agroinfiltraron hojas de *N. benthamiana* con una suspensión de *A. tumefaciens* conteniendo la construcción y se observó en microscopio confocal (Zeiss LSM 780 Multifoton) dos días después de la agroinfiltración. La imagen de la izquierda muestra la superposición de fluorescencias de GFP (*VTE3.1::GFP*) en color verde y la clorofila en color rojo debido a su auto-florescencia, mientras que el panel de la derecha muestra la superposición de la fluorescencia de GFP (*VTE3.2::GFP*) y clorofila. La barra blanca representa 10µm. La fotografía de microscopía del panel A fue adaptada de Quadrana 2014 y la del panel derecho B fue gentilmente cedida por Juliana Almeida (datos no publicados).

A partir de la disponibilidad del genoma de alta calidad de tomate (The Tomato Genome Consortium, 2012), fue posible la identificación de un gen parálogo a *VTE3(1)*, nombrado *VTE3(2)*. Este tiene una longitud casi tres veces mayor que su parálogo (6.060 pb respecto de 2.803 pb, respectivamente), se encuentra localizado en el brazo norte del cromosoma 3 y codifica para una proteína de tamaño similar a *VTE3(1)* con una similitud superior al 77%. Por otro lado, la comparación a nivel de secuencia de aminoácidos entre los productos de los alelos salvajes de *S. pennelli* y el del cultivar *S. lycopersicum* presentó valores aún mayores al 98% para ambos parálogos. Las dos proteínas poseen dominios de metil transferasa, varios motivos de unión a S-adenosilmetionina y un péptido señal de direccionamiento a cloroplasto en su extremo amino terminal (Figura 22B, C y D). Predicciones basadas en análisis *in silico* que evalúan la potencial funcionalidad de proteínas de interés a través de la identificación de mutaciones disruptivas (<http://sift.jcvi.org/>) no reveló ningún residuo comprometido.

La caracterización de esta proteína mediante la búsqueda de ortólogos en diferentes especies reveló que la misma sólo se encuentra presente en organismos fotosintéticos. Sin embargo la síntesis de tococromanoles ha sido reportada en *Plasmodium falciparum* un organismo no fotosintético que causa la malaria y sintetiza tocoferoles durante su estadio intra-eritrocitario (Sussaman y col. 2011, Cassera y col. 2004). La búsqueda de secuencias similares a *VTE3* en el genoma de este parásito, no identificó ningún ortólogo, debido a su lejana filogenia o a la ausencia de la misma. Al realizar el alineamiento de las secuencias y un análisis de agrupamiento por Neighbor joining se encontró que al igual que en tomate (y otras *Solanaceas*) el gen *VTE3* está duplicado en especies de Vitales y Malpighiales (Figura 24). Sin embargo tanto algas como Monocotiledóneas y varias familias de Dicotiledóneas (como Brassicaceas y Fabales) poseen sólo una copia del gen, indicando que la presencia de una copia simple es el carácter más primitivo. Resulta interesante resaltar que no existe sintenia entre las regiones genómicas conteniendo ambas copias del gen *VTE3* en los cromosomas 3 y 9, sugiriendo que el mecanismo que originó ambas copias del gen fue una duplicación génica simple y no el producto de un evento de poliploidización ni duplicación genómica total o parcial, como aquellas que forjaron la arquitectura genómica de las diferentes familias de plantas vasculares (Jiao y col. 2011) incluyendo al tomate (The Tomato Genome Consortium, 2012). Esta observación es respaldada por la estimación del tiempo de divergencia entre los parálogos. El cálculo de sustituciones por sitio sinónimo (Ks) entre *VTE3(1)* y *VTE3(2)* (método de Nei-Gojobori) resultó en $1,33 \pm 0,21$ sustituciones por sitio sinónimo entre ambos parálogos. Teniendo en cuenta que en el genoma de tomate se han identificado dos eventos de duplicación genómica, uno antiguo en el rango de 1,8 a 2,5 mutaciones por sitio sinónimo (poliploidización i, The Tomato Genome Consortium, 2012) y otro más reciente y exclusivo de

Solanáceas entre 0,8 y 1,0 sustituciones por sitio sinónimo (Song y col. 2012), la duplicación de VTE3 resulta independiente de estos eventos masivos de ampliación genómica. Con el fin de estimar la edad de la duplicación de VTE3, se utilizó un ritmo de $6.1 \cdot 10^{-9}$ sustituciones sinónimas por sitio por año (Lynch y Conery 2003) y se calculó que el tiempo transcurrido desde la duplicación es de 109 ± 18 millones de años, luego de la separación entre Monocotiledóneas y Dicotiledóneas (anterior a 125 millones de años atrás). No obstante, no es posible asegurar con estos resultados que este evento de duplicación haya ocurrido previo a la radiación entre Rósidas y Astéridas (y una de las duplicaciones se perdió luego en Brassicaceas y Fabales) o bien posterior e independiente en Astéridas (Solanaceas), Vitales (ej. *Vitis vinifera* L.) y Malpigiales (ej. *Populus trichocarpa*).

El análisis de datos de contenidos de tocoferoles en frutos de tomate públicamente disponibles (Tomato Funtional Genomics Database:

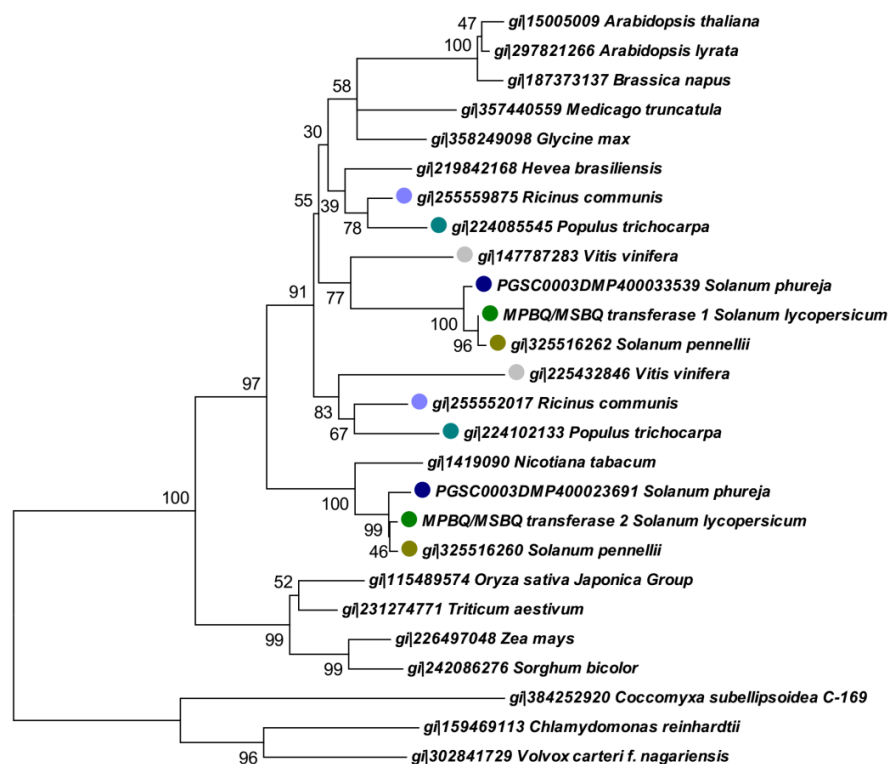


Figura 24. Árbol filogenético de VTE3. A partir de la lista de proteínas ortólogas de otras especies de plantas y algas obtenidas por BLASTP en la base de datos de NCBI, se realizó un análisis filogenético con el método de Neighbor Joining (Bootstrap=10.000) incluido en el programa MEGA5. Las plantas que presentaban más de una proteína VTE3, producto de la traducción de genes duplicados, se identificaron con un punto del mismo color a la izquierda del nombre. Adaptado de Quadrana 2013.

<http://ted.bti.cornell.edu/>) reveló la presencia de otro QTL para estos compuestos en el cromosoma 3. En este sentido, la IL3.1 muestra incrementos de 155,5% y 126% en comparación con el genotipo parental donante cultivar M82 para α - y tocoferol total, respectivamente, mientras que las formas δ - y γ -

tocoferol se encuentran reducidas en 40.7% y 66.7%. Asimismo, estudios previos del grupo mostraron que la expresión del gen *VTE3(2)* en diferentes tejidos de tomate correlaciona con la acumulación de α - y tocoferol total como así también con otros genes que codifican para enzimas involucradas en la biosíntesis de tococromanoles (Quadrana y col, 2013), proponiendo a *VTE3(2)* como un nodo regulatorio del metabolismo de tococromanoles en la red de co-expresión de los genes centrales de esta vía.

Un análisis de la estructura génica de este parólogo revela que al igual que lo observado en *VTE3(1)*, *VTE3(2)* posee una inserción de un elemento transponible de tipo retrotransposon SINE en su región promotora. A su vez, existen dos alelos para este elemento, existiendo 86% de identidad de secuencia entre los alelos de *S. lycopersicum* y *S. pennellii*. Este elemento se encuentra físicamente dentro de una región diferencialmente metilada del genoma de tomate (DMR09944) reportada en la base de datos de Tomato Epigenome Database (<http://ted.bti.cornell.edu/epigenome/>) (Figura 22A). Estas evidencias permiten postular que la regulación transcripcional de parólogo *VTE3(2)* también podría involucrar mecanismos epigenéticos.

En el marco de la segunda de las hipótesis planteadas en el capítulo introductorio de esta tesis: **“La biosíntesis de tococromanoles es regulada diferencialmente a nivel de las enzimas involucradas y los genes *VTE3(1)* y *VTE3(2)* interactúan, determinando en parte los contenidos de estos compuestos en plantas de tomate”**, se presentan a continuación los resultados de experimentos realizados para abordar los siguientes objetivos:

- 1) Construir una población F2 derivada del cruzamiento de las líneas de introgresión IL3.1 e IL9.2.6.1 portando todas las posibles combinaciones de alelos salvajes (*S. pennellii*) y cultivados (*S. lycopersium*) de ambos parálogos *VTE3(1)* y *VTE3(2)*.
- 2) Analizar el efecto fenotípico sobre los individuos recombinantes F2 de la población arriba mencionada, desafiando a la población en dos ambientes con condiciones contrastantes: a campo y en invernáculo.
- 3) Analizar los patrones de metilación de las regiones regulatorias y los niveles de expresión de estos genes en los individuos recombinantes F2 de la población mencionada en las mismas condiciones ambientes contrastantes.

III.2.Resultados

III.2.1 Obtención de una población F2 recombinante para los alelos silvestres y cultivados de los genes parálogos VTE3(1) y VTE3(2)

Con el fin de estudiar posibles interacciones alélicas y epialélicas entre los genes *VTE3(1)* y *VTE3(2)*, generamos una población segregante derivada del cruzamiento de una IL (IL9.2.6.1) conteniendo el alelo silvestre de *VTE3(1)* en el fondo genético de la especie domesticada (*S. lycopersicum lycopersicum* cv M82) con otra IL (IL3.1) conteniendo el alelo silvestre de *VTE3(2)* en el mismo fondo genético. Las plantas F1 resultantes fueron autofecundadas para dar origen a la población F2 de 201 individuos que segregan los alelos silvestres y cultivados de los genes *VTE3(1)* y *VTE3(2)*. Para identificar la combinación alélica de estos genes en cada uno de los individuos F2 resultantes, se llevó a cabo un análisis de marcadores moleculares. En particular, estos marcadores fueron especialmente seleccionados por ser codominantes, es decir que permiten diferenciar genotipos homocigotas y heterocigotas, por estar físicamente sobre los genes de interés (Figura 25). En el capítulo V (sección V.2.3 Desarrollo y aplicación de marcadores moleculares para el genotipado de la población recombinante F2 Cruzamiento IL 3.1 x IL9.2.6.1) se encuentra una descripción más detallada de la población recombinante F2, tanto para su obtención como de los análisis moleculares que permitieron el genotipado de los individuos que la componen. En total fueron genotipadas con ambos marcadores 80 plantas de la población F2, las cuales no mostraron distorsión en la segregación alélica de ambos genes (Prueba $\chi^2 = 6,35$; el valor umbral de χ^2 para 8 grados de libertad para un $p < 0,01$ es 20,1). Luego, 3 individuos F2 independientes para cada genotipo fueron seleccionados para los siguientes análisis. Estos 27 individuos abarcan el espectro completo de posibles combinaciones alélicas (ver más adelante, figura 29).

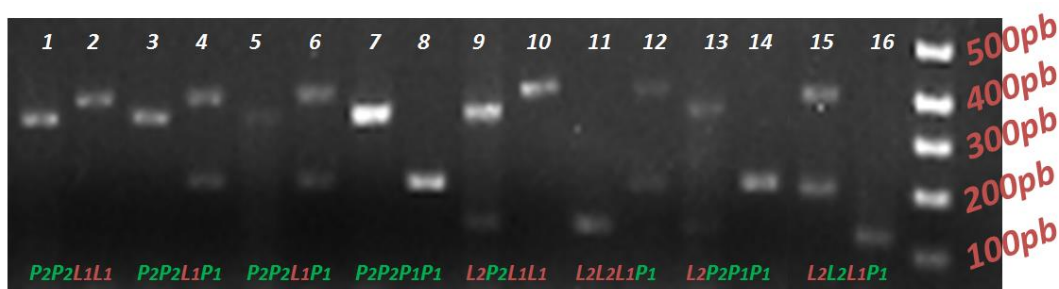


Figura 25. Genotipado de 8 plantas con distintas combinaciones de primers para los genes parálogos *VTE3* de las 80 recombinantes F2 (cruzamiento IL9.2.6.1 x IL3.1) con dos marcadores moleculares ligados a *VTE3(1)* y *VTE3(2)*. Gel de agarosa 1 % (bromuro de etidio) el cual muestra en las 16 primeras calles los productos de PCR de 8 plantas F2. El marcador *VTE3(2)* que permite discriminar entre los alelos silvestres de *S. pennelli* (indicado como P_2 en verde) y el cultivar *S. lycopersicum* cv M82 (indicado como L_2 en rojo) están sembrados en las calles impares mientras los respectivos productos para *VTE3(1)* ambos alelos *S. pennelli* (representados como P_1 en verde) y *S. lycopersicum* (indicados como L_1 en rojo) están sembrados en las calles pares. El marcador de peso molecular se encuentra en la última calle y en rojo sus respectivos pesos moleculares.

III.2.2 Fenotipado y análisis de expresión y metilación del ADN de los parentales de las líneas F2 recombinantes, desafiados en dos ambientes contrastantes

Mediante la técnica de HPLC (ver sección V.3 Medición de datos fenotípicos (contenido de tococromanoles) por HPLC), se cuantificaron los contenidos de α - β - γ - δ -tocoferol y tocoferol total en hojas y frutos de los parentales de la población presentada precedentemente. Consistente con los resultados de Quadrana y col (2013), se observaron mayores niveles de tocoferoles en hojas en comparación con los frutos, lo cual está en la línea con la función propuesta para estas moléculas en la protección del aparato fotosintético en cloroplastos (Krieger-Liszkay y Trebst 2006). Asimismo observamos que α -tocoferol es la forma más abundante tanto en frutos como en hojas, validando observaciones previas (Abushita y col. 1997, Moco y col. 2007).

Tanto los dos genotipos parentales (IL9.2.6.1 y IL3.1) como el padre recurrente (cv. M82) de la población interespecífica original (Eshed y Zamir, 1995) mostraron alta variación en los contenidos de tocoferoles, especialmente cuando se comparan plantas cultivadas en condiciones ambientales contrastantes. También notamos que el efecto del ambiente sobre los contenidos totales de tocoferol es homogéneo para todas las líneas parentales, sugiriendo un efecto marginal de la interacción genotipo x ambiente. Las tres líneas parentales mostraron incrementos en los contenidos medios totales de tocoferoles en fruto de las plantas crecidas en condiciones de campo (en la localidad de Chacras de Coria, Mendoza) respecto de los contenidos medidos en los mismos genotipos en condiciones controladas de invernadero. Similarmente, en las hojas se observó un incremento de los contenidos de tocoferol total de 112,2% en las plantas cultivadas en condiciones de campo respecto de aquellas provenientes de invernaderos. Estos resultados se observan en el gráfico de barras particionado para cada una de los genotipos: IL9.2.6.1 e IL3.1 y cv. M82 (Figura 26).

Con el objetivo de investigar los mecanismos moleculares responsables de los cambios de tocoferol observados, se evaluaron los niveles de ARNm de los genes *VTE3(1)* y *VTE3(2)*, como así también la metilación del ADN de sus regiones promotoras. Basados en resultados previos del grupo que indican la existencia de regulación diferencial de ambos alelos de *VTE3(1)*, diseñamos *primers* específicos (Figura 22 A) para ambos alelos, lo cual permite diferenciar tanto la expresión como la metilación en las reacciones de PCR y qPCR. De esta manera, las mediciones de expresión y metilación son alelo-específicas. Los niveles de expresión de *VTE3(1)* y *VTE3(2)* varían significativamente entre genotipos, indicado un alto componente genético en la regulación transcripcional de estos genes. Notamos que, en frutos maduros, los niveles de ARNm de *VTE3(1)* son mayores que los de *VTE3(2)* para ambos alelos, tanto

en las líneas parentales como en el padre recurrente cv M82. Por otro lado, *VTE3(2)* presenta mayores niveles de ARNm en las hojas fuente (Figura 27). Este comportamiento se acentúa aún más en tejidos de plantas cultivadas a

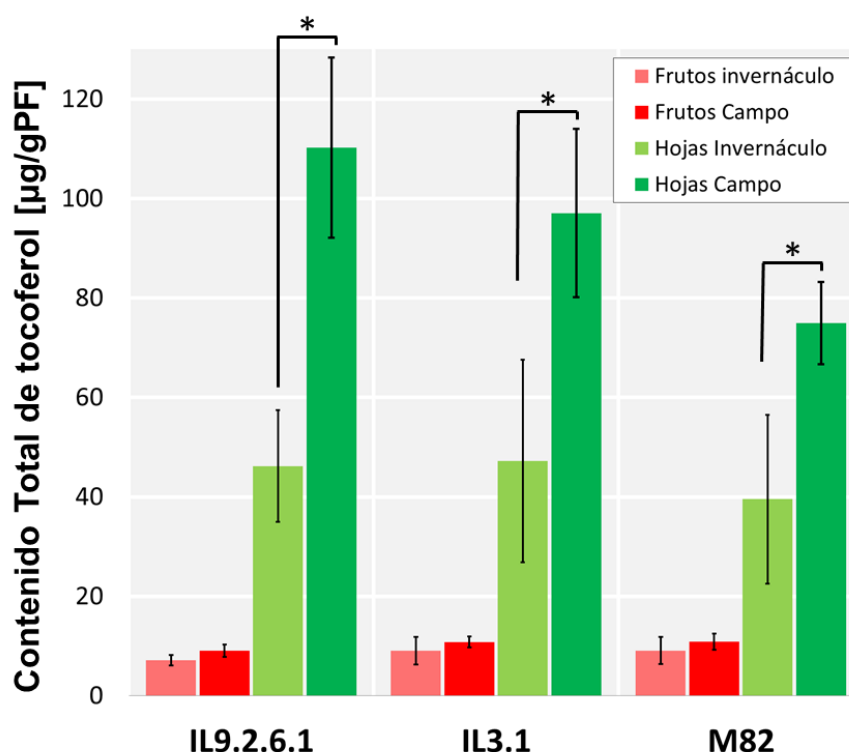


Figura 26. Contenidos de tocoferoles totales de frutos y hojas en condiciones contrastantes para los parentales de la F2; IL9.2.6.2, IL3.1 y para el padre recurrente cv M82. El gráfico de barras muestra la modulación de los contenidos de tocoferol totales por parte del ambiente para los frutos (barras rojas) y las hojas (barras verdes) de las mismas plantas clonales parentales de la F2 (IL9.2.6.2, IL3.1 y M82) cosechadas en dos ambientes contrastantes. Las barras rojas y verdes claras indican que los frutos y hojas han sido cosechados en invernáculo mientras que las barras de color oscuro, indican condiciones de campo. Los * indican diferencias significativas entre condiciones para un mismo parental y tejido (p-valor <0.05 prueba T-student).

campo en comparación a condiciones controladas de invernáculo. Por lo tanto, podemos proponer la existencia de una regulación transcripcional tejido-específica diferencial de ambos parálogos (Figura 27). Esta divergencia funcional ya ha sido reportada en otros genes codificantes de enzimas involucradas en la biosíntesis de tococromanoles, como es el caso de *GGPS*, sugiriendo que este mecanismo de regulación podría ser recurrente para este metabolismo (Quadrana y col. 2013). Adicionalmente observamos que los niveles de metilación del ADN en las regiones promotoras de estos genes dependen fuertemente del ambiente y el genotipo, (Figuras 27 y 28). En conjunto, estos resultados indican que la regulación de la expresión de *VTE3(1)* y *VTE3(2)* es sensible a la interacción del genotipo x ambiente. Por otra parte el alelo *S. lycopersicum* de *VTE3(1)* presenta los mayores niveles de ARNm en los frutos maduros de las plantas del cultivar recurrente cv. M82 cultivadas en condiciones de campo. Un comportamiento semejante había sido observado en

experimentos previos del laboratorio estudiando los niveles de expresión de este gen en condiciones de campo y en condiciones controladas de invernáculo (Quadrana y col. 2014). Estos resultados indican que la respuesta transcripcional al ambiente es robusta y de alta heredabilidad. Dado que esta respuesta transcripcional al ambiente del gen *VTE3(1)* en *S. lycopersicum* cv. M82 ha sido asociada a cambios epigenéticos (Quadrana et al 2014), medimos los niveles de metilación de los promotores de *VTE3(1)* y *VTE3(2)* (Figura 28). Observamos que los niveles de metilación del promotor correlacionan negativamente con los niveles expresión de *VTE3(1)* ($p=-0,71$; $p<0.01$). Es decir, la alta y baja expresión se asocia, respectivamente, a una baja y alta metilación de su promotor (Figuras 27 y 28). Por otro lado, detectamos muy baja expresión del alelo domesticado de *VTE3(2)* en frutos maduros, consistente con una alta metilación de su promotor. A su vez, cuando analizamos hojas fuentes, donde *VTE3(2)* es altamente expresado, observamos bajos niveles de metilación de su promotor como así también una mayor expresión en plantas cultivadas a campo (Figuras 27 y 28).

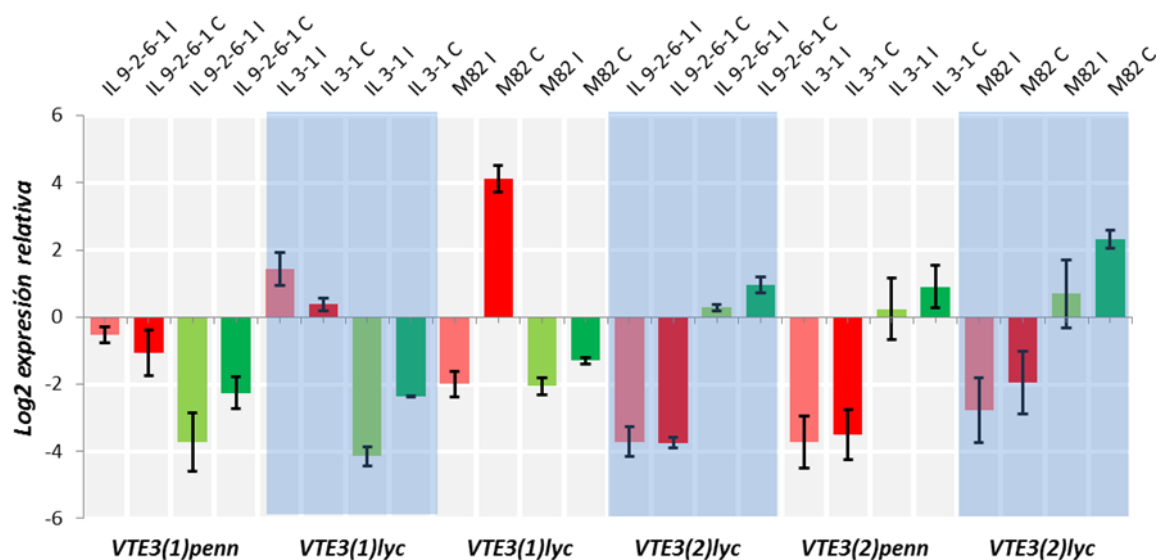


Figura 27. Gráfico de barras para niveles de expresión relativa de los ARNm de los genes *VTE3(1)* y *VTE3(2)* en condiciones contrastantes para los parentales de la F2, IL9.2.6.2, IL3.1 y el padre recurrente cv M82 en tejido de frutos y hojas. El gráfico de barras muestra la modulación de la expresión relativa para cada uno de los alelos de *VTE3(1)* y *VTE3(2)* en dos condiciones contrastantes de las mismas plantas clonales expresada en función del Log_2 (más detalle de las RT-QPCR y el cálculo de la misma se encuentra en las secciones V.4.1 y V.4.2, en materiales y métodos). Cada una de las barras corresponde a un parental, el cual se encuentra indicado en el eje superior de abscisas seguido de una letra I para la condición de plantas cosechadas en invernáculo para los frutos, indicados con barras rojas claras y para las hojas con barras verdes claras. Mientras que la letra C representa a las plantas cosechadas a campo representadas por barras rojas oscuras y las hojas en verde oscuras. Los paneles claros y oscuros alternados de izquierda a derecha, indican la los alelos correspondientes a las barras para cada uno de los parentales (los mismos están referidos en el eje de abscisas inferior representados como el gen *VTE3* correspondiente seguido de *lyc*, para *S. lycopersicum* y *penn* para *S. pennelli* respectivamente). Los * indican diferencias significativas entre tejidos o entre ambientes para un mismo parental y alelo (p -valor <0.05 prueba T-student). Los datos están relativizados al promedio de expresión de los dos genes para ambos alelos y posteriormente transformados a Log_2 .

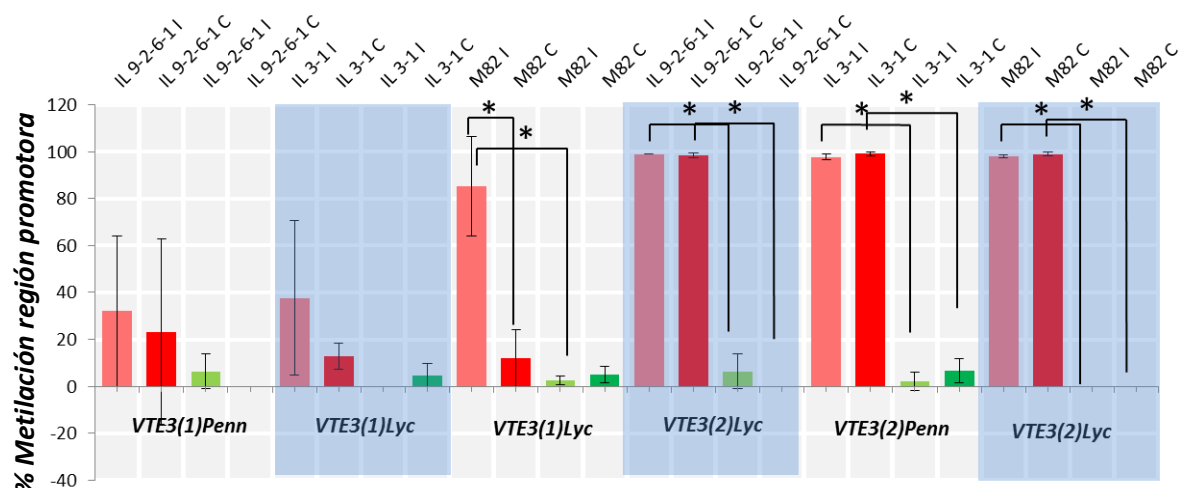


Figura 28. Gráfico de barras para niveles de metilación de la región promotora de los genes *VTE3(1)* y *VTE3(2)* en tejidos de fruto y hoja, de los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en condiciones contrastantes de invernáculo y campo. El gráfico de barras muestra la modulación por parte del ambiente sobre los niveles de metilación de la región promotora para cada uno de los alelos de *VTE3(1)* y *VTE3(2)* para las mismas plantas clonales parentales de la F2: IL9.2.6.2, IL3.1 y M82, expresados como % de metilación (calculado a partir del análisis con la enzima endonucleasa MCrBC, ver sección V.5 para más detalle). Cada una de las barras corresponde a un parental, el cual se encuentra indicado en el eje superior de abscisas seguido de una letra I para la condición de plantas cosechadas en invernáculo para los frutos, indicados con barras rojas claras y para las hojas con barras verdes claras. Mientras que la letra C representa a las plantas cosechadas a campo representadas por barras rojas oscuras y las hojas en verde oscuras. Los paneles claros y oscuros alternados de izquierda a derecha indican la los alelos correspondientes a las barras para cada uno de los parentales (los mismos están referidos en el eje de abscisas inferior representados como el gen *VTE3* correspondiente seguido de *lyc*, para *S. lycopersicum* y *penn* para *S. pennelli* respectivamente). Los * indican diferencias significativas entre tejidos o entre ambientes para un mismo parental y alelo (p-valor <0.05 prueba T-student).

Estos resultados apoyan la hipótesis de que la regulación de la expresión de estos genes, tanto tejido específica como ambiente específica, tiene un importante componente epigenético. Con el fin de evaluar el impacto de esta regulación transcripcional en los contenidos de tocoferoles, comparamos la acumulación de α + γ respecto de β + δ tocoferol (Figura 29). La lógica detrás de esta comparación es que las enzimas codificadas por los genes *VTE3s* determinan el flujo metabólico entre estas formas de tocoferoles, favoreciendo la acumulación de α + γ tocoferoles. Esta relación en hojas es marcadamente mayor en las tres líneas parentales cultivadas en condiciones controladas en relación a condiciones de campo. Estos incrementos se deben a cambios significativos en la acumulación de las isoformas α -y β -tocopherol (Figura 29A y B). Por otro lado, en los frutos provenientes de plantas cultivadas en condiciones de campo esta relación decrece para la línea IL9.2.6.1, principalmente debido a diferencias significativas en β -tocopherol, y se mantiene relativamente constante en los genotipos IL3.1 y M82. No obstante, las tres líneas parentales presentan un incremento promedio del 22% de α + γ

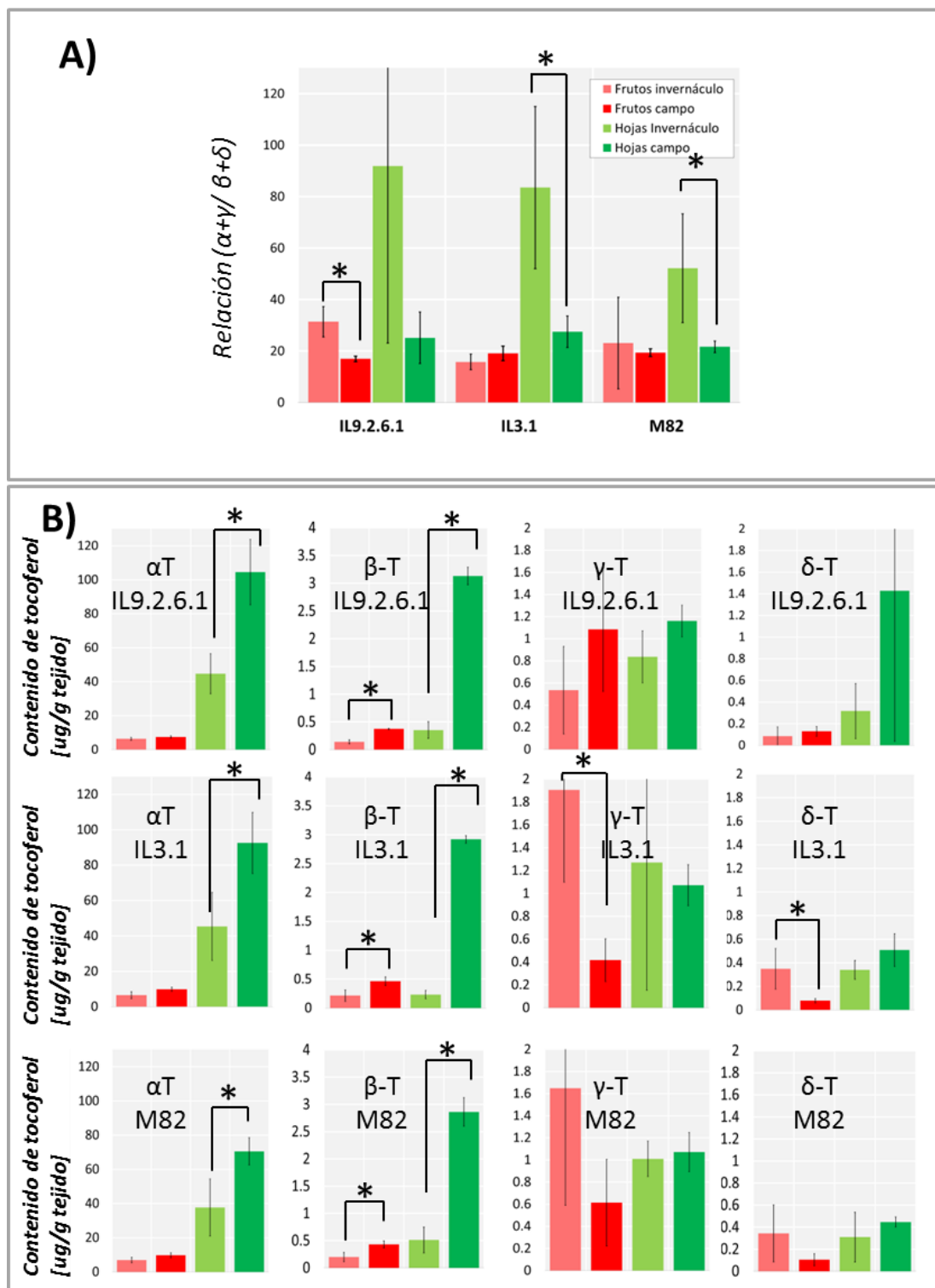


Figura 29. Relación de la acumulación de (α+γ/ β+δ) en frutos y hojas en condiciones contrastantes para los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en tejido de fruto y hoja. En el panel A) El gráfico de barras indica la relación promedio (α+γ/ β+δ) calculada a partir de 3 plantas parentales de la F2 IL9.2.6.2, IL3.1 y M82 para frutos y hojas en invernáculo, frutos (barras rosadas) y hojas (barras verde claro) Seguido de izquierda a derecha por frutos cosechados en condiciones de campo (barras rojas oscuras) y hojas (verdes oscuras). En el panel B) se indican los contenidos de las 4 isoformas de tocoferol α-, β-, γ-, δ-tocoferol de izquierda a derecha en ese mismo orden tanto para frutos como en hojas para los parentales: IL9.2.6.1; IL3.1 y M82 en el siguiente orden de arriba hacia abajo. * Indica el diferencias significativas entre condiciones para un mismo genotipo parental y tejido (p<0.05 prueba t-student).

tocoferoles en frutos de plantas cultivadas en condiciones de campo frente a los de plantas cultivadas en condiciones de invernáculo mientras que los niveles de β - y δ -tocoferol aumentan y disminuyen respectivamente, enmascarando un posible aumento en las formas dependientes de VTE3. Resultados similares fueron reportados por Tang y col. (2016) en plantas transgénicas de lechuga sobreexpresando el ortólogo de *A. thaliana* *AtVTE3*, las cuales tampoco presentaron decrecimiento en la acumulación β - y δ -tocoferoles de acuerdo a lo esperado y observado en plantas transgénicas de *AtVTE3* en soja (Van Eenennaam y col. 2003) y en nuestro laboratorio en frutos de tomates silenciando *VTE3* por VIGS (Quadrana y col. 2014). Estos resultados sugieren que otros factores, además de la expresión en frutos de los genes *VTE3s*, podrían estar operando en la regulación del metabolismo de tocoferoles, incluyendo la actividad de otras enzimas río abajo en la vía como VTE4 y VTE1. Por ejemplo, los cambios de β - y δ -tocoferol podrían ser explicados por un aumento en el flujo a través de VTE4 (Figura 21, capítulo II). Contar con datos de expresión de *VTE4* y *VTE1* en estos materiales permitiría avanzar en la demostración de esta hipótesis.

A continuación caracterizamos la expresión y metilación de los genes *VTE3s* en la IL9.2.6.1, la cual es homocigota para el alelo silvestre de *VTE3(1)*. En frutos maduros, la expresión del alelo silvestre del gen *VTE3(1)* es mayor que la del gen *VTE3(2)* (Figura 27) como era de esperar dada la expresión tejido-específica de estos genes (Quadrana y col, 2013). Adicionalmente, ninguno de los dos genes mostró variaciones significativas entre plantas cultivadas a campo o en condiciones controladas (Figura 27), sugiriendo que la expresión en fruto del alelo silvestre de *VTE3(1)* es insensible a los ambientes estudiados. Consistente con este resultado, los niveles de metilación del promotor son relativamente bajos e invariables (Figura 28). No obstante la expresión estable de ambos genes en frutos, la relación $\alpha+\gamma/\beta+\delta$ es mayor en frutos de plantas cultivadas en condiciones controladas (Figura 29). Sugiriendo que otros factores, además de la expresión en frutos de los genes *VTE3s*, estarían operando en la regulación del metabolismo de VTE. Análisis de expresión en hojas muestran que los niveles de mRNA de *VTE3(2)* son predominantes e incrementados en plantas cultivadas a campo (Figura 27). Tanto en plantas cultivadas en condiciones controladas como a campo, los niveles de metilación de sus regiones promotoras son bajos y estables (Figura 28). Similar a lo observado en frutos, la relación ($\alpha+\gamma/\beta+\delta$) en hojas tiende a disminuir en condiciones de campo, a pesar de mostrar alta variabilidad intragenotipo, impidiendo detectar diferencias estadísticamente significativas (Figura 29).

Similar al análisis conducido para la línea IL9.2.6.1, evaluamos la expresión y metilación de los genes *VTE3s* en frutos de la línea IL3.1, la cual posee el alelo silvestre del gen *VTE3(2)*. Consistente con el resto de los genotipos analizados,

la expresión en frutos de *VTE3(2)* es muy baja en ambos ambientes (Figura 27). Inesperadamente, los niveles de ARNm de *VTE3(1)* tienden a ser mayores en plantas cultivadas en condiciones controladas comparado a plantas cultivadas a campo, aunque estas diferencias no son estadísticamente significativas (Figura 27). No obstante estos resultados, la relación $\alpha+\gamma/\beta+\delta$ en los frutos no presenta diferencias en plantas cultivadas en condiciones de campo en comparación con aquellas cultivadas en invernáculo (figura 29). Finalmente, la expresión en hojas del alelo silvestre de *VTE3(2)* es más elevada comparado con el alelo domesticado *VTE3(1)*. La expresión de este último es inducida en hojas de plantas cultivadas a campo relativo a las cultivadas en invernáculo. Sin embargo, la relación $(\alpha+\gamma/\beta+\delta)$ presenta reducciones significativas entre muestras de plantas cultivadas a campo frente a las cosechadas en invernáculo (figura 29). Estos resultados sugieren que tanto la regulación transcripcional como su impacto en la acumulación del producto de las enzimas codificadas por estos genes son más complejas de lo propuesto previamente. A su vez, podrían indicar la existencia de reguladores en *trans* de *VTE3(1)*, probablemente localizados en la región introgresada de la IL3.1.

III.2.3. Caracterización fenotípica y molecular de individuos de la población F2.

Como se explica en la sección *III.2.1*, a partir del genotipado de los individuos de la población recombinante F2 se identificaron entre 3 y 19 individuos para cada uno de los 9 genotipos recombinantes posibles. En particular, 1/4 de las líneas recombinantes portan la misma constitución alélica para los dos genes *VTE3s* que las líneas parentales IL9.2.6.1, IL3.1 y que el parental recurrente cv M82 (Figura 30). Por tanto, estos individuos reconstituyen los genotipos parentales y son de gran utilidad para estudiar la estabilidad tanto de los fenotipos de interés como de los mecanismos moleculares subyacentes. El análisis por marcadores moleculares nos permitió identificar al grupo de recombinantes R7 como la reconstitución del genotipo control cv M82, mientras que los R9 y R1 representan la reconstitución de los genotipos parentales IL3.1 e IL.9.2.6.1, respectivamente.

En primer lugar analizamos los contenidos de tocoferoles de estas líneas cultivadas tanto en condiciones de invernáculo como en condiciones de campo, comparándolos con los datos fenotípicos obtenidos para los genotipos parentales. La relación $(\alpha+\gamma/\beta+\delta)$ fue similar entre las líneas parentales y sus respectivos genotipos F2 reconstituidos (Figura 31) en ambos tejidos, con la excepción de los frutos de plantas de la IL3.1 cultivados en condiciones controladas de invernáculo que mostraron una menor relación $(\alpha+\gamma/\beta+\delta)$ que su respectivo genotipo reconstituido R9 (Figura 31). Estos resultados sugieren

que la relación entre tocoferoles es considerablemente estable luego de los cruzamientos realizados para obtener la población F2.

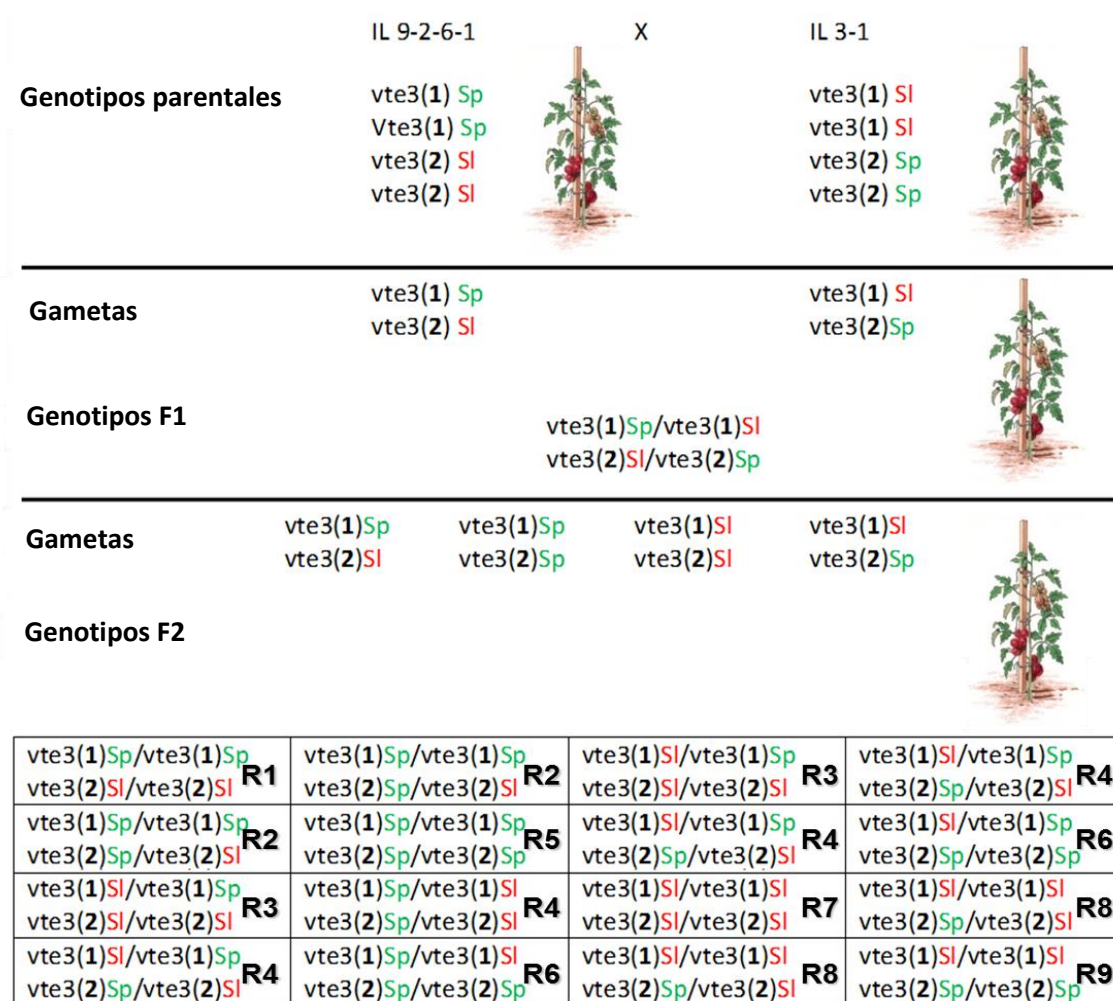


Figura 30. Esquema de obtención de los 9 genotipos recombinantes. Los posibles genotipos recombinantes F2 están indicados como R1, R2,...,R9. Estos fueron generados a partir del cruce de las líneas de introgresión IL9.2.6.1 y IL3.1 obteniendo la F1 y finalmente la F2. Los alelos correspondientes a la especie *S. lycopersicum* (SI) y *S. Pennelli* (Sp) tanto para el gen *VTE3(1)* como para *VTE3(2)* están indicados en verde y rojo respectivamente más detalles de la construcción de esta población y su segregación ver secciones V2.1-V2.3

No obstante la estabilidad observada a nivel de los metabolitos, la expresión del alelo *S. lycopersicum* de los genes *VTE3(1)* y *VTE3(2)* en frutos de plantas R7 cultivadas a campo mostraron disminuciones significativas comparadas al parental M82 (Figura 32). Asimismo, se detectó una reducción significativa en la expresión del alelo *S. lycopersicum* de *VTE3(1)* en los frutos de la línea R9 cultivada en el invernáculo en comparación con el parental IL3.1 cultivado en el mismo ambiente (Figura 32A). Aún más notorios fueron los cambios observados en la expresión de ambos alelos de los dos genes en las hojas: en las líneas R1 y R9 cultivadas en condiciones de invernáculo se observó un

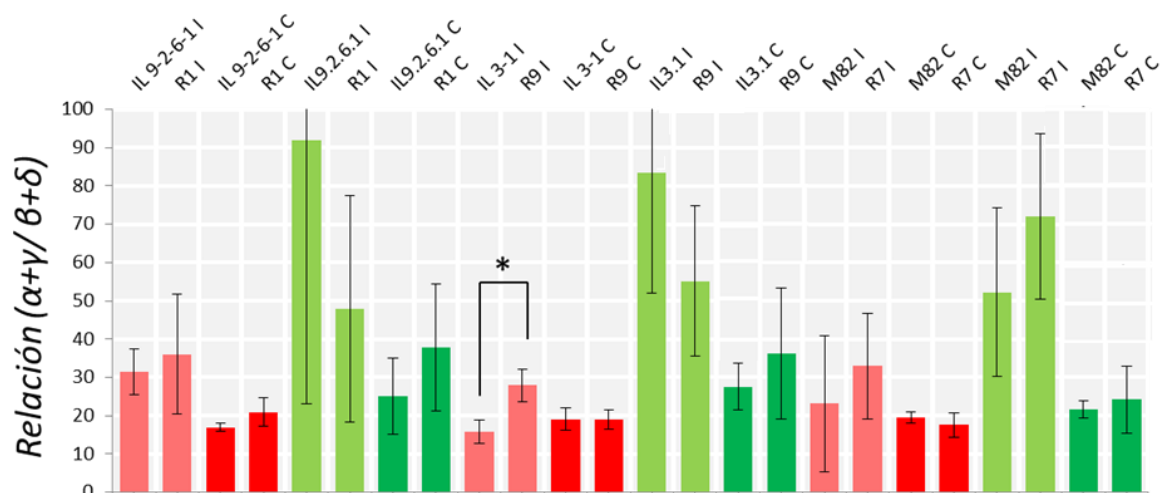


Figura 31. Relación de la acumulación de ($\alpha+\gamma/\beta+\delta$) en frutos en condiciones contrastantes de invernáculo y campo para los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en comparación a sus genotipos reconstituyentes R1, R9 y R7 respectivamente. La relación ($\alpha+\gamma/\beta+\delta$) fue calculada para 3 plantas correspondiente a cada uno de los grupos parentales y reconstituyentes estudiados definidos por los marcadores ligados a *VTE3(1)* y *VTE3(2)*. Los respectivos genotipos están indicados en el eje de abscisas superior seguidos de una letra I, para la condición de plantas cultivadas en condiciones de invernáculo y una letra C, para el tratamiento a campo. Las barras rojas claras y oscuras representan la relación para frutos cosechados en invernáculo y campo respectivamente mientras que las barras verde claras y oscuras representan la relación para frutos cosechados en invernáculo y campo respectivamente. * Diferencias significativas entre el parental y su reconstituyente están representadas con $p<0.05$ prueba t-student.

incremento en los niveles de expresión de los alelos silvestres y cultivados, respectivamente, en relación a los niveles observados en los correspondientes parentales IL9.2.6.1 e IL3.1. En cambio, en la línea R7 el alelo de *lyc* de ambos genes mostró una disminución de los niveles de ARNm respecto de aquellos del parental correspondiente (cv M82) en las hojas de las plantas cultivadas en condiciones de invernáculo. El alelo silvestre de *VTE3(2)* también mostró una disminución en los niveles de ARNm en la línea R9 respecto del parental IL3.1, pero esto ocurrió exclusivamente en las hojas de plantas cultivadas en condiciones de campo (Figura 32B).

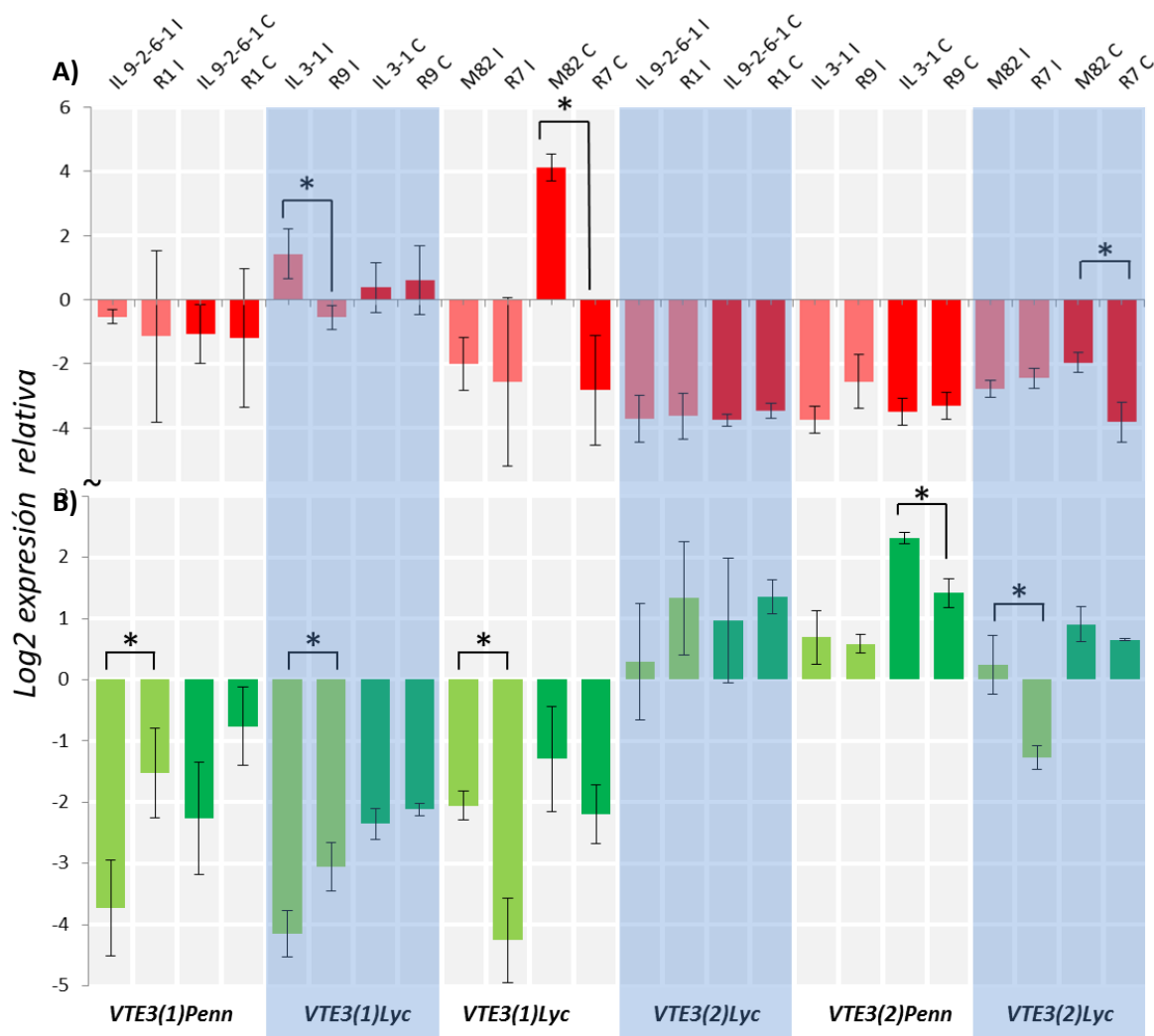


Figura 32. Niveles de expresión en frutos y hojas en condiciones contrastantes de invernáculo y campo para los parentales de la F2, IL9.2.6.2, IL3.1 y M82 en comparación a sus genotipos reconstituyentes R1, R9 y R7 respectivamente. El gráfico de barras muestra la modulación de la expresión relativa para cada uno de los alelos de VTE3(1) y VTE3(2) en dos condiciones contrastantes mismas plantas clonales parentales de la F2: IL9.2.6.2, IL3.1 y M82 *versus* sus respectivas reconstituyentes recombinantes R1, R9 y R7 en hojas panel **A**) y en frutos panel **B**). Los datos están expresados en función del Log₂ (más detalle de las RT-QPCR y el cálculo de la misma se encuentra en las secciones V.4.1 y V.4.2, en materiales y métodos). Cada una de las barras corresponde a un parental o reconstituyente, el cual se encuentra indicado en el eje superior de abscisas seguido de una letra I para la condición de plantas cosechadas en invernáculo (Castelar) para los frutos, indicados con barras rojas claras y para las hojas con barras verdes claras. Mientras que la letra C representa a las plantas cosechadas a campo en Mendoza representadas por barras rojas oscuras y las hojas en verde oscuras. Los paneles claros y oscuros alternados de izquierda a derecha, indican la los alelos correspondientes a las barras para cada uno de los parentales (los mismos están referidos en el eje de abscisas inferior representados como el gen VTE3 correspondiente seguido de *lyc*, para *S. lycopersicum* y *penn* para *S. pennelli* respectivamente). Los * indican diferencias significativas entre parentales y sus reconstituyentes para un mismo ambiente (p-valor < 0.05 prueba T-student).

Con todo, se vuelve a observar una regulación transcripcional tejido-especifica diferencial de ambos parálogos. Además, es evidente que la producción de la población segregante F2, ya sea durante las recombinaciones ocurridas en las

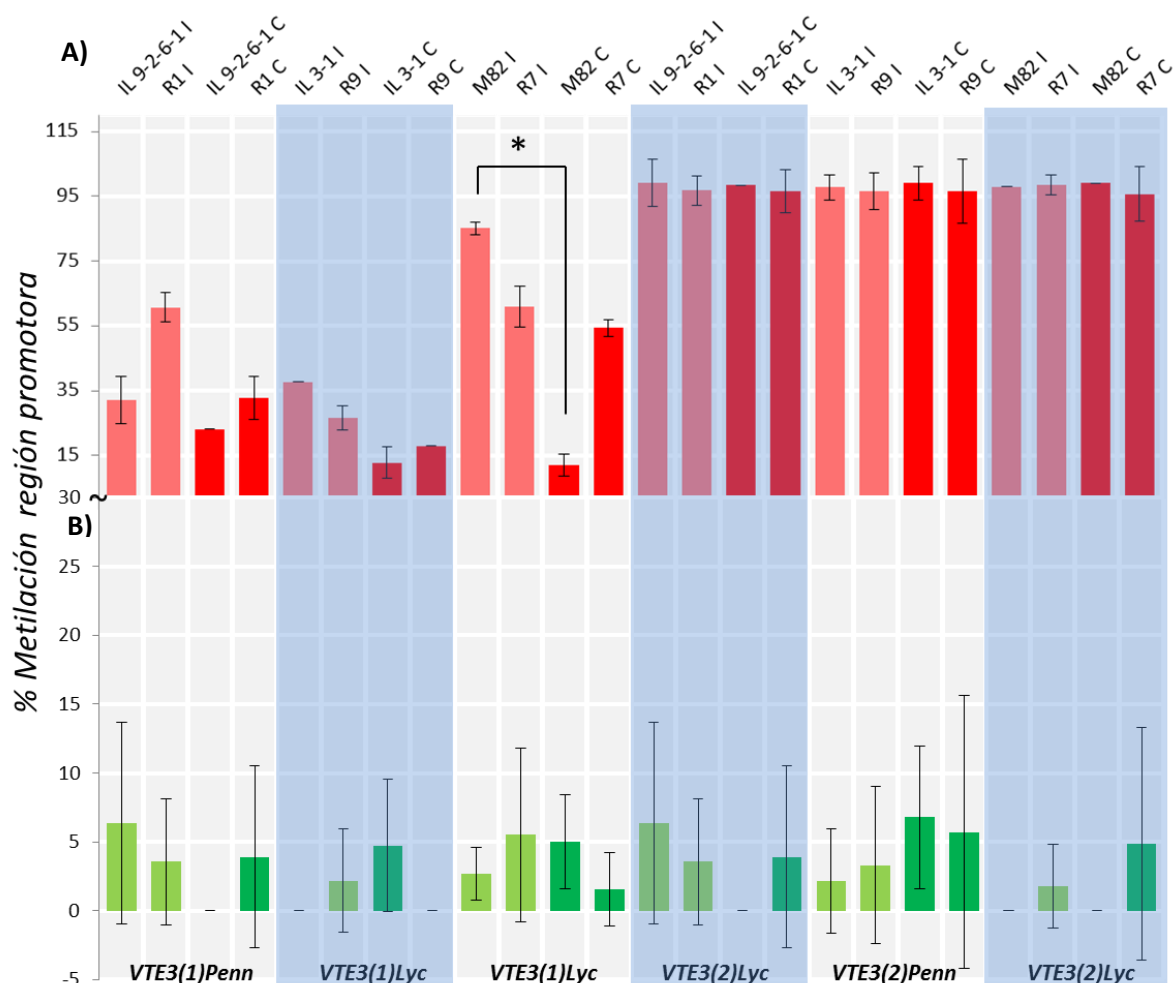


Figura 33. Gráfico de barras para niveles de metilación para la región promotora de los genes *VTE3(1)* y *VTE3(2)* en tejido de fruto y hoja, de los parentales de la F2, IL9.2.6.2, IL3.1 y M82 campo en comparación a sus genotipos reconstituyentes R1, R9 y R7 respectivamente, en condiciones contrastantes de invernáculo y campo. El gráfico de barras muestra la modulación por parte del ambiente sobre los niveles de metilación de la región promotora para cada uno de los alelos de *VTE3(1)* y *VTE3(2)* para las mismas plantas clonales parentales de la F2: IL9.2.6.2, IL3.1 y M82, versus sus respectivas reconstituyentes recombinantes R1, R9 y R7. Los datos están expresados como % de metilación (calculado a partir del análisis con la enzima endonucleasa MCrBC, ver sección V.5 para más detalle). Cada una de las barras corresponde a un parental o su recombinante, el cual se encuentra indicado en el eje superior de abscisas seguido de una letra I para la condición de plantas cosechadas en invernáculo (Castelar) para los frutos, indicados con barras rojas claras y para las hojas con barras verdes claras. Mientras que la letra C representa a las plantas cosechadas a campo en Mendoza representadas por barras rojas oscuras y las hojas en verde oscuras. Los paneles claros y oscuros alternados de izquierda a derecha, indican la los alelos correspondientes a las barras para cada uno de los parentales (los mismos referidos en el eje de abscisas inferior representados como el gen *VTE3* correspondiente seguido de *lyc*, para *S. lycopersicum* y *penn* para *S. pennellii* respectivamente). Los * indican diferencias significativas entre entre ambientes para un mismo parental y alelo (p-valor <0.05 prueba T-student).

dos meiosis que dieron origen a las líneas recombinantes o simplemente a las dos generaciones extras que posee la población F2, desencadenó alteraciones de los niveles de ARNm de los genes *VTE3s*. Los niveles de metilación de las regiones regulatorias de estos genes mostraron patrones bien diferentes entre tejidos; mientras en las hojas ambos alelos de los dos genes alcanzaron valores menores del 15 % de metilación, en los frutos los niveles de metilación difieren marcadamente entre *VTE3(1)* y *VTE3(2)*. Consistente con el bajo nivel de expresión de este último en los frutos (Figura 33), la región regulatoria analizada muestra niveles de metilación cercanos al 100 % en este tejido (Figura 33A). En cambio *VTE3(1)* muestra niveles de metilación de su región promotora que varían desde 10 % (en frutos del cv M82 cultivado en condiciones de campo) hasta cerca del 80 % (en frutos de este mismo genotipo pero en condiciones de invernáculo). Consistente con lo que se observa en los niveles de expresión, la línea R7 presenta menores niveles de metilación que el parental M82 en los frutos del invernáculo y lo opuesto ocurre en condiciones de campo (Figura 33A). Los resultados descritos precedentemente sugieren que la inestabilidad a nivel transcripcional, y en menor medida en la relación $(\alpha+\gamma/\beta+\delta)$ de tocoferoles, podría ser debida a factores epigenéticos en *trans* establecidos en respuesta al cruzamiento entre los parentales IL3.1 y IL9.2.6.1 y/o durante las meiosis entre los cromosomas homólogos en estado homocigota primero (líneas parentales) y heterocigotas luego (F1) que originaron la población F2 analizada en esta tesis. Alternativamente, la variación entre los genotipos parentales y sus respectivos genotipos reconstituidos en la población F2 podría deberse simplemente a variaciones intergeneracionales o la presencia de introgresiones de origen desconocido en los parentales que segregan independientemente de *VTE3(1)* y *VTE3(2)*.

III.2.4 Análisis de la población F2 y evaluación de la contribución de factores genéticos, epigenéticos y ambientales como así también sus interacciones en los niveles de tocoferol y expresión de VTE3(1) y VTE3(2)

El estudio de los parentales utilizados para generar la población F2 (sección III.2.2) permitió identificar interacciones entre el genotipo y el ambiente como así también una importante asociación entre la expresión de los genes *VTE3s* y los niveles de metilación de sus secuencias promotoras. Adicionalmente, la caracterización molecular de líneas de la población F2 con genotipos parentales reconstituyentes sugirió la existencia de interacciones entre ambos parálogos. Con el objetivo de ampliar nuestro conocimiento sobre los factores que contribuyen a la variación en los niveles de tocoferol y la expresión y metilación de *VTE3(1)* y *VTE3(2)*, extendimos nuestro análisis a la población F2 para las nueve combinaciones alélicas posibles para los genes *VTE3s* a partir de 27 plantas representativas del *pool* de 80 plantas F2 en ambas condiciones (Figura 30). Para esto, es preciso implementar pruebas estadísticas que sean capaces de cuantificar y evaluar no solo la contribución

de cada factor a la variabilidad de la población, sino también de la interacción entre estos factores. Los modelos lineales mixtos (MLM) han sido propuestos como los más indicados para este tipo de análisis (Moore y col. 2018, Douglas y col. 2015). MLM es un modelo paramétrico apoyado en varios supuestos, siendo la distribución normal de los datos un requisito importante. Al inspeccionar las mediciones notamos una fuerte desviación en la normalidad en todas las variables analizadas como por ejemplo en la acumulación de δ -tocoferol en tejido de hojas (Figura 34). Con el objetivo de normalizar los datos se ensayaron diferentes transformaciones, incluyendo transformación logarítmica (Log_{10}), raíz cuadrada (Sqrt) y BoxCox, las cuales fueron evaluadas por una prueba de normalidad de Shapiro-Wilk. Entre las transformaciones ensayadas, BoxCox logró normalizar el mayor número de variables y por lo tanto fue implementada en los análisis subsiguiente (Tabla 11).

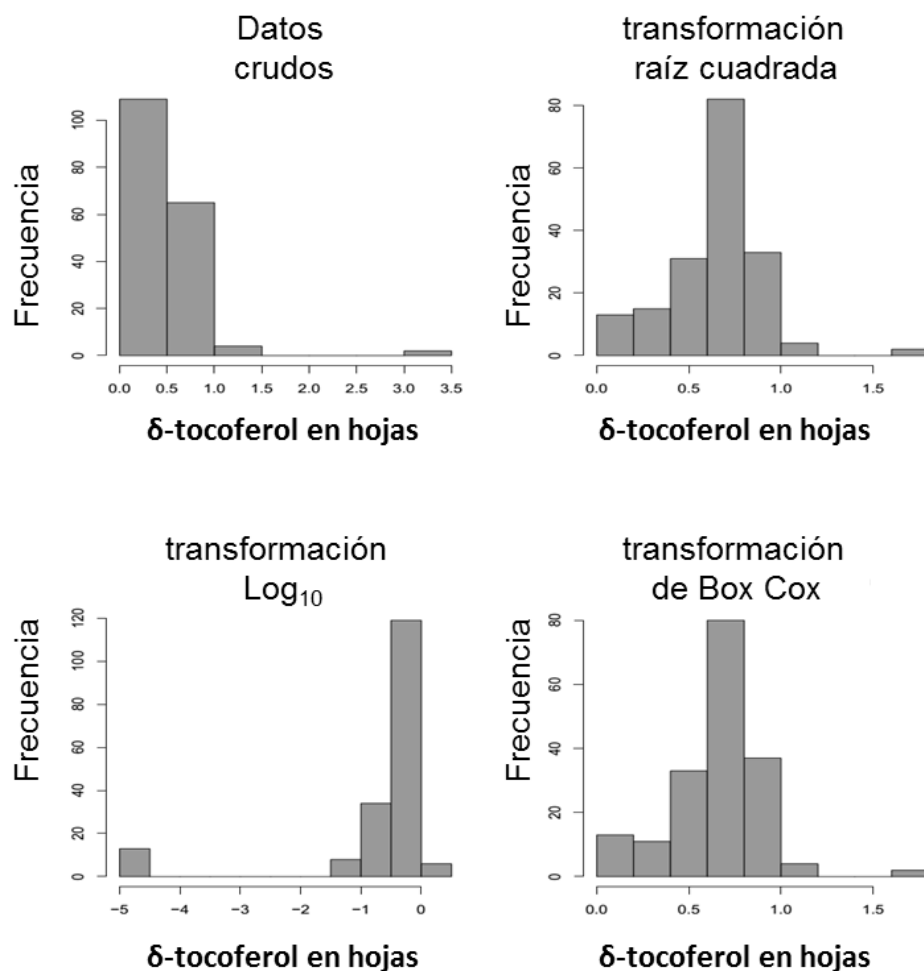


Figura 34. Gráficos de distribuciones de frecuencias para las distintas transformaciones ensayadas sobre los datos de δ -tocoferol en hojas de la F2 para los datos de δ -tocoferol de las 27 plantas clonales F2. La distribución de frecuencias para 27 plantas F2 con cada una de las combinaciones de alelos para *VTE3(1)* y *VTE3(2)* en ambos ambientes presentaron distribuciones lejos de la normalidad.

Tabla 11 Resumen de la prueba de Shapiro-Wilks para todas las variables analizadas antes y después de aplicar la transformación BoxCox.

Variables	Lambda Box Cox	p-valor Shapiro Wilks pre-transformación	p-valor Shapiro Wilks post-transformacion
α -tocoferol en frutos	1.03	0.152	0.157
β -tocoferol en frutos	0.46	4.1 E-8	5.58E-08
γ -tocoferol en frutos	0.22	4.13E-06	0.001
δ -tocoferol en frutos	0.26	2.33E-15	0.001
Tocoferol total en hojas	1.01	0.0119	0.02
α -tocoferol en hojas	0.26	1.42E-07	1.51E-06
β -tocoferol en hojas	0.3	5.93E-07	1.12E-06
γ -tocoferol en hojas	-0.2	2.40E-11	0.002
δ -tocoferol en hojas	0.46	2.20E-16	2.61E-09
Tocoferol total en hojas	0.26	0.13E-05	1.73E-06
Exp. VTE3(1) S. lyc frutos	1.03	0.18E-05	0.157
Exp. VTE3(1) S. penn frutos	0.46	4.1 E-8	4.38E-05
Exp. VTE3(2) S. lyc frutos	0.23	4.13E-06	0.001
Exp. VTE3(2) S. penn frutos	0.22	2.33E-15	0.001
Exp. VTE3(1) S. lyc hojas	1.01	0.28E-05	0.001
Exp. VTE3(1) S. penn hojas	0.26	1.42E-07	2.51E-05
Exp. VTE3(2) S. lyc hojas	0.3	5.93E-07	1.11E-02
Exp. VTE3(2) S. penn hojas	-0.2	2.40E-11	0.001
Met. VTE3(1) S. lyc frutos	0.26	0.14E-05	1.73E-06
Met.VTE3(1) S. penn frutos	1.02	0.34E-05	0.157
Met.VTE3(2) S. lyc frutos	0.50	4.1 E-8	5.58E-08
Met.VTE3(2) S. penn frutos	0.48	5.120E-16	2.61E-09
Met.VTE3(1) S. lyc hojas	0.26	0.13E-05	1.73E-06
Met.VTE3(1) S. penn hojas	1.03	0.24E-05	0.157
Met.VTE3(2) S. lyc hojas	0.48	4.1 E-8	5.58E-08
Met.VTE3(2) S. penn hojas	0.50	2.30E-16	2.61E-09
Prueba de Shapiro wilks para las variables del modelo MLM antes y después de la transformación de Box Cox. Exp y Met indican expresión y metilación respectivamente para cada uno de los genes VTE3s. El valor de Lambda de la función Box-cox indica el factor o función por el cual se transforman los datos de la respectiva variable. $p < 0,05$ no presentan distribuciones normales, mientras que valores con $p > 0,05$ si lo son.			

Para la determinación del mejor modelo se tuvieron en cuenta los coeficientes AIC y BIC (para mas detalle de estos coeficientes de información Akaike (AIC) y criterio de información bayesiano (BIC) ver materiales y métodos sección V.X). El modelo MLM aplicado aquí para el cálculo del análisis de la varianza (ANOVA) asociada a cada uno de los isómeros de tocoferol y tocoferol total, tanto en hojas como en frutos esta descrito por los siguientes factores: *genotipo o grupo genético*, definido por cada una de las nueve combinaciones de los alelos de *S. lycopersicum* y *S. pennelli* los genes *VTE3(1)* y *VTE3(2)* posibles dentro de la F2 el cual se encuentra equilibrado, con tres plantas por

genotipo evitando la sobrerrepresentación por parte de alguno de los alelos *VTE3s*; el *ambiente* (campo o invernáculo), la *interacción genotipo x ambiente*; el efecto de cada *planta* recombinante individual, el cual contempla la aquellas recombinaciones diferenciales entre plantas dentro del grupo genético y finalmente la *expresión y metilación relativas* de cada uno de los alelos *VTE3s* como variables independientes del modelo.

En frutos maduros el modelo MLM aplicado fue significativo (p -valor <0.001) explicando una gran proporción de la varianza y su heterogeneidad entre las cuatro isoformas de tocoferol. El residuo (o varianza no explicada) fue generalmente acotado salvo para la acumulación de γ -tocopherol en la cual representa 42% de la varianza total (Figura 35). Interesantemente el porcentaje de la varianza asociada al nivel de metilación de las regiones promotoras llega a representar entre 41,9 y 75,6% de la varianza asociada a β - y δ -tocopherol respectivamente, presentándose como la variable con mayor efecto sobre la acumulación de las isomorfos α -, β - y δ - tocoferol. En segundo orden se destaca el factor ambiente el cual compone entre 23,4% y 28,4 % de la acumulación de las isoformas γ - y α -tocopherol, respectivamente. Por otro lado, la expresión del alelo *S. lycopersicum* de *VTE3(1)* explica entre 2,1 y 12,8% de la varianza total de α - y β -tocopherol. No obstante, ambos alelos *VTE3(1)* son preponderantes en cuanto a su rol en este tejido; sumados llegan a explicar hasta 75,6% de la acumulación de δ -tocopherol en frutos. Estos resultados son congruentes con la regulación transcripcional tejido-especifica diferencial de ambos parálogos para las plantas reconstituyentes R1, R7 y R9 observada previamente. Por otro lado, la interacción genotipo x ambiente es máxima para la isomorfos γ - y β -tocopherol con un rango de entre 17,3% y 5,2%. Sin embargo, la combinación de alelos que define al genotipo de cada una de los 27 individuos aquí evaluados, representa solamente entre 0,9 y 9,8% de la variación de δ - y α -tocopherol. La variable planta, la cual tiene en cuenta las recombinaciones individuales, explica mayor proporción con un efecto de entre 2,9 y 19.2% δ - y β -tocopherol. La partición de la varianza para α -tocopherol fue similar que aquella obtenida para tocoferol total, consistente con esta forma representando el 85% del tocoferol total.

Al igual que en frutos maduros, todos los modelos de acumulación de tocoferol en hojas fueron significativos (p -valor <0.001) con una distribución de la varianza aun mas heterogénea entre las diferentes formas de tocoferol (Figura 36). Esto sugiere un amplio grado de respuesta dependiendo del isómero y la variable que evaluada. Por ejemplo, más de la mitad de la varianza de los contenidos de β -tocopherol (53,9%) se explica por el ambiente, mientras que para la acumulación de γ -tocopherol este representa tan solo un 3%. Sorprendentemente en hojas el efecto del nivel de metilación de las regiones promotoras de los genes *VTE3s*, considerando este como la suma del efecto

de los cuatro alelos al igual que en fruto también contribuyen fuertemente al carácter con 31,6, 18,9 y 9,2% del total de la varianza para las isoformas α - γ - β -tocoferol, respectivamente mientras que para δ -tocoferol, no presenta efecto alguno (Figura 36). Contrariamente, esta isoforma es explicada en gran parte por la expresión del alelo silvestre de *VTE3(1)* el cual compone 49,9% de la varianza fenotípica. Cuando evaluamos la expresión global, es decir la sumatoria de la expresión de los dos alelos, esta explico entre 17,2 y 77,7% para α - y γ - tocoferol, respectivamente. Tanto el genotipo como la variable planta presentaron un pobre poder explicativo en estos modelos. El primero solo compone 3,1% del total de la varianza de α -tocoferol mientras que para las restantes isoformas no presenta efecto alguno, al igual que ocurre con la variable planta. Esto podría deberse a un “enmascaramiento” por otra variable del modelo, como la expresión de los alelos, u otros factores no analizados. Al igual que en frutos la distribuciones de las proporciones de la varianza en las hoja son semejantes entre α -tocoferol y tocoferol total.

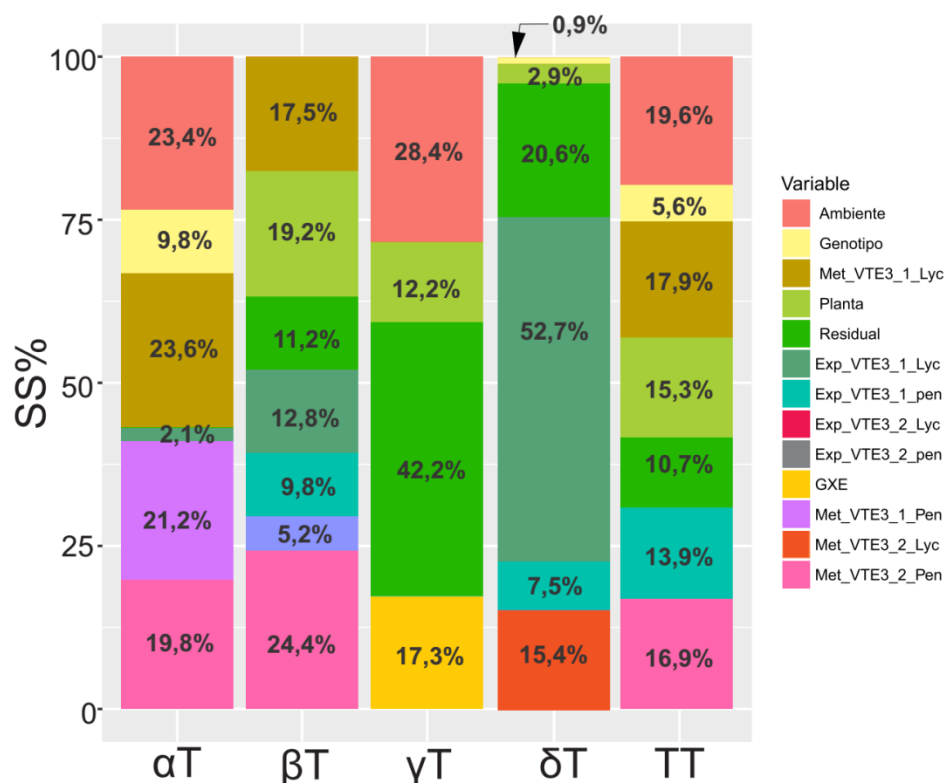


Figura 35 Disección de la varianza fenotípica total para las distintas isoformas de tocoferol en frutos para los datos de toda la F2 cruza (IL9.2.6.1x IL3.1). Para cada carácter fenotípico la figura presenta la proporción de cada uno de los efectos en la suma de cuadrados %SS de la varianza total para cada una de las isoformas de tocoferol representadas como α -T; β -T; γ -T; δ -T; TT para alfa, beta, gama ; delta y tocoferol total respectivamente. Los factores del modelo o variables se encuentran descriptos con sus respectivos colores a la izquierda de la figura. La expresión de los 4 alelos se indica como Exp_ seguido del gen correspondiente *VTE3(1)* o *VTE3(2)*; la metilación de cada uno de los correspondientes alelos fue indicada como Met_ seguido de el gen correspondiente *VTE3(1)* o *VTE3(2)*; Pen y Lyc indican los alelos *S. pennelli* y *S. lycopersicum* respectivamente.

En conjunto, estos resultados muestran que la exploración de la varianza fenotípica mediante los modelos MLM en ambos tejidos es explicada en gran medida por el *ambiente* como así también por el nivel de *metilación* de las regiones promotoras y la *expresión* de los genes *VTE3s*. Con el objetivo de cuantificar la relación de estas tres variables entre ellas se recurrió nuevamente a los modelos MLM, definiendo las variables metilación y expresión como la suma individual de las contribuciones de cada uno de los alelos *VTE3s*. A continuación, de forma similar al análisis sobre los caracteres fenotípicos, se presenta una evaluación de estos modelos sobre los datos de expresión y metilación en ambos ambientes sobre los 27 individuos F2 analizados.

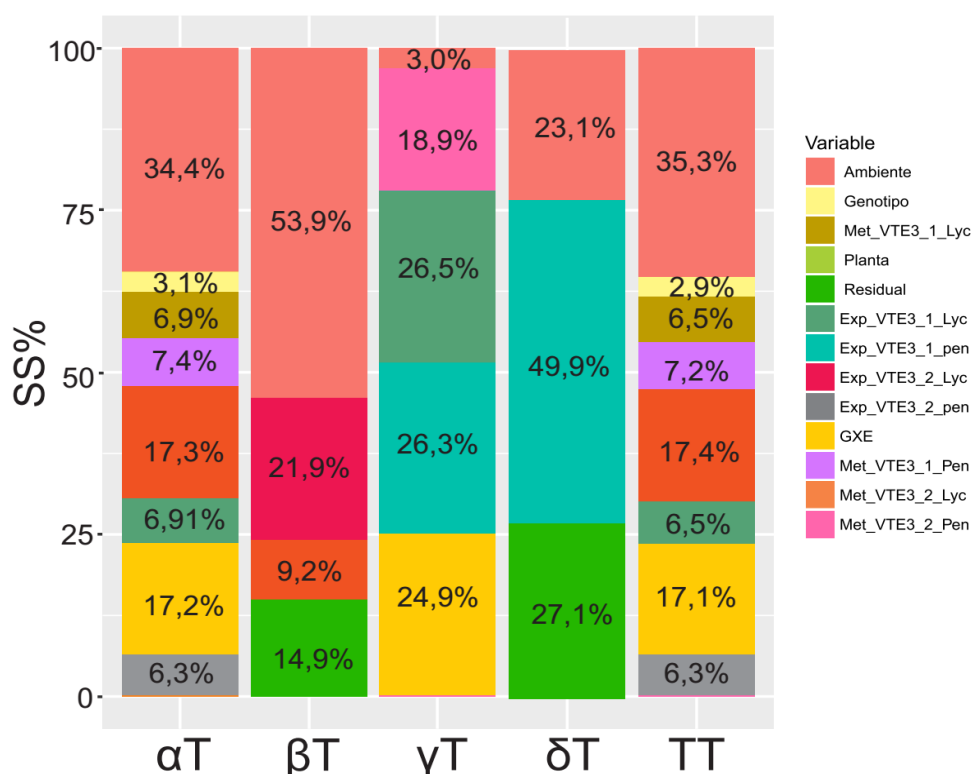


Figura 36. Disección de la varianza fenotípica total para las distintas isoformas de tocoferol en hojas para los datos de toda la F2 cruza (IL9.2.6.1x IL3.1). Para cada carácter fenotípico la figura presenta la proporción de cada uno de los efectos en la suma de cuadrados %SS de la varianza total para cada una de las isoformas de tocoferol representadas como α-T; β-T; γ-T; δ-T; TT para alfa, beta, gama ; delta y tocoferol total respectivamente. Los factores del modelo o variables se encuentran descriptos con sus respectivos colores a la izquierda de la figura. La expresión de los 4 alelos se indica como Exp_ seguido del gen correspondiente *VTE3(1)* o *VTE3(2)*; la metilación de cada uno de los correspondientes alelos fue indicada como Met_ seguido de el gen correspondiente *VTE3(1)* o *VTE3(2)*; Pen y Lyc indican los alelos *S. pennelli* y *S. lycopersicum* respectivamente.

El modelo MLM definido para la expresión incluyó como factores las variables: *planta*, *metilación*, *genotipo* y *ambiente*. El mismo fue significativo ($p < 0.001$). los porcentajes de la varianza en la expresión fueron explicados por la *metilación* y *genotipo* de manera similar entre hojas y frutos (Figura 37). La

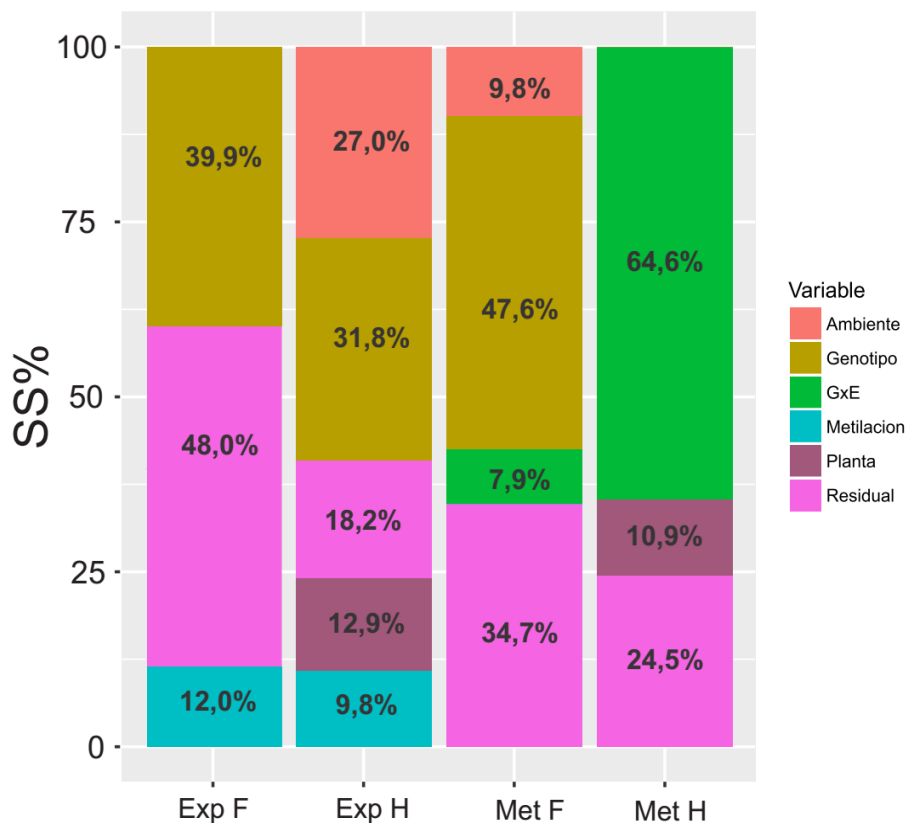


Figura 37. Disección de la varianza asociada a la Expresión (Exp) y Metilación (Met) en tejidos de hojas (H) y frutos (F) a partir de modelos MLM para cada una de las variables teniendo en cuenta como factores : el ambiente, genotipo, GXE, metilación y planta para toda la población F2 cruza (IL9.2.6.1x IL3.1). Cada uno de los porcentajes de la figura representa la proporción o efecto en la suma de cuadrados %SS total para el modelado de la expresión (Exp) y metilación (Met) de las 27 plantas F2 en tejidos de hojas (H) y frutos (F), tomando como factores del modelo: el ambiente, genotipo, GXE, metilación y planta definido como el efecto de cada una de las plantas individuales.

metilación explica 12,04% y 9,83% de la varianza en la expresión en frutos y hojas, respectivamente, mientras que el *genotipo* tiene mucho mayor efecto sobre el nivel de expresión de los genes (39,9% y 31,8% de la varianza en frutos y hojas respectivamente). Estos resultados coinciden con lo observado previamente en las secciones II.2.2 y II.2.3, donde el genotipo, definido como la combinación de alelos *VTE3s*, coordina la expresión individual de cada alelo sobre el resto tanto para los parentales IL9.2.6.1, IL3.1 y cv M82 como para las líneas recombinantes F2 con los genotipos reconstituídos. Contrariamente, el *ambiente* como variable explicativa de la *expresión* de los genes *VTEs* se diferencia fuertemente entre tejidos, contribuyendo 27,1% de la varianza en hojas pero 0% en frutos. Al igual que como se señaló en la disección de la varianza de los caracteres fenotípicos de tocoferol, esto podría deberse a un “enmascaramiento” o pseudorreplicación (Hurlbert 1984) del *genotipo* dentro de otra/s de las variables del modelo, inflando indirectamente el poder explicativo de la metilación en este caso. Cabe remarcar el alto porcentaje de varianza

residual (48% y 18,2% en frutos y hojas, respectivamente), el cual indica la existencia de otros factores no contemplados en el modelo.

En contraste con lo observado para la expresión, el modelado de la *metilación* (Figura 37) expuso una gran contribución por parte del ambiente en ambos tejidos. Su varianza en frutos está explicada en un 9,8% por el ambiente y la interacción genotipo x ambiente comprende 7,9% de la variabilidad observada. En hojas la interacción genotipo x ambiente fue preponderante explicando la mayor proporción de la varianza de la metilación (64,6%). Este resultado permite explicar el efecto no observado por ambiente sobre la varianza de la expresión, la cual soporta nuestra hipótesis de que la variable *metilación* se podría estar enmascarando indirectamente. Sorprendentemente el *genotipo* tuvo un bajo aporte sobre el nivel de metilación en las hojas, mientras que en frutos este llegó a explicar 47,6% de la variabilidad. No obstante, este podría estar explicado a través de la contribución de la variable planta indirectamente, la cual aporta 10,9% de la varianza total de la metilación, sugiriendo que este efecto tejido-específico de las recombinaciones diferenciales entre plantas individuales de un mismo grupo genético F2, poseen mayor peso que la contribución del grupo genético.

En conjunto la disección de la varianza de la metilación y expresión para los genes *VTE3(1)* y *VTE3(2)* de población F2 recombinante sugiere que la combinación de alelos *VTE3s*, definida aquí como genotipo, sería la variable de mayor importancia con consecuencias directas sobre la expresión y la metilación para ambos tejidos (Figuras 37). Por otro lado el efecto del ambiente sobre la expresión ocurriría indirectamente a través de la metilación como variable (Figura 37).

Con el objetivo de caracterizar mejor la interacción entre variables, condujimos múltiples análisis de correlación pareadas entre todas las variables analizadas (Figura 38). Interesantemente los coeficientes de correlación para la expresión entre los alelos de *S. pennelli* y *S. lycopersicum VTE3(1)* ($\delta:0,73$; $p<0,001$) y *VTE3(2)* ($\delta:0,72$; $p<0,001$) son muy elevados en hojas (Figura 38, indicados como H1). Así mismo observamos una fuerte correlación entre la expresión de los parálogos silvestres de *VTE3(1)* y *VTE3(2)* ($\delta:0,90$; $p<0,001$), soportando los resultados del MLM que indicaban una coordinación/interacción entre estos dos genes. No obstante, en frutos esto no se observa en los niveles de mensajeros o metilación de la región promotora entre genes *VTE3s*, los coeficientes son de moderados a bajos. Las únicas correlaciones identificadas ocurren entre el nivel de metilación de la región promotora del parálogo silvestre *VTE3(2)* sobre la expresión de *VTE3(1)* *S. pennelli* (F1: $\delta:0,64$; $p<0,001$) y entre los niveles de metilación y expresión de *VTE3(1)* *S. lycopersicum* correlacionan inversamente en este tejido (F2: $\delta:-0,62$; $p<0,001$).

Coeficiente de correlación de Spearman

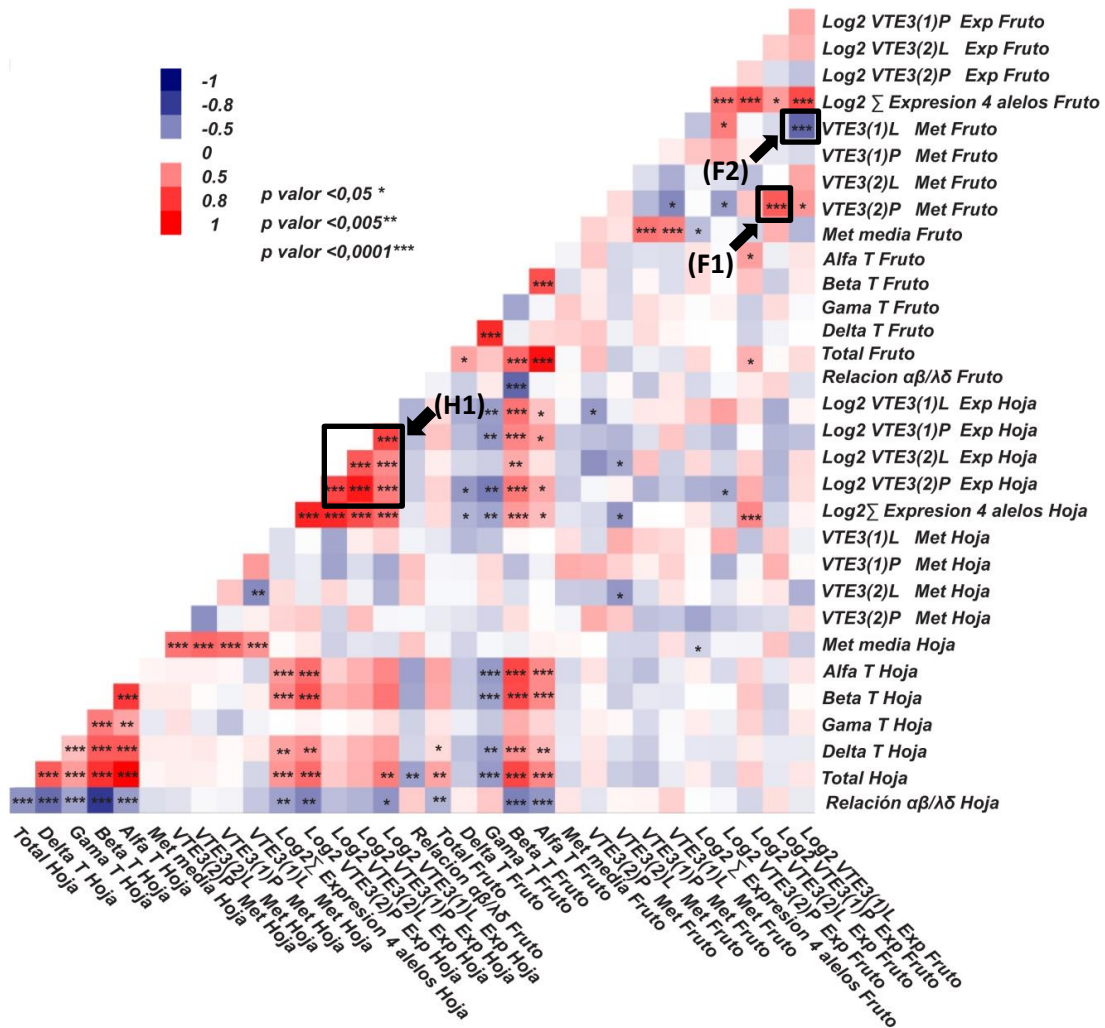


Figura 38. Gráfico de *heatmap* de correlaciones de Spearman entre variables fenotípicas de tocoferol, niveles relativos de expresión y metilación de los genes *VTE3s* así como de la sumatoria de la expresión de los cuatro alelos y su promedio de metilación para tejidos de hojas y frutos independientemente del ambiente. Las correlaciones más relevantes se encuentran indicadas como H1 para hojas y F1 y F2 para frutos enmarcados por un rectángulo negro. La escala de colores del *heatmap* entre las distintas variables del modelo se encuentra en el vértice superior izquierdo, cada color indica valores entre -1 y 1 para el correspondiente coeficiente de Spearman, correlaciones positivas en rojo y negativas en azul. El grado de significancia de los mismos está indicado con asteriscos; $p < 0,05^*$; $p < 0,005^{**}$; $p < 0,0001^{***}$.

III.3 Discusión

Los resultados del análisis genético presentado en el capítulo II expusieron que la heredabilidad del carácter estudiado en esta tesis es relativamente baja. Esto había sido reportado previamente en tomate (Schauer y col, 2008; Sauvage y col. 2014). Sumado a esto, los valores de heredabilidad para el tocoferol total se reducen drásticamente cuando se analizan datos de plantas cultivadas en condiciones de estrés hídrico (0,45 en condiciones de cultivo control comparado con 0,239 en condiciones estresantes). Estos resultados son consistentes con la alta sensibilidad al ambiente de este carácter en plantas de olivo frente a condiciones ambientales de déficit hídrico e incremento de la temperatura, las cuales afectaron la expresión de varios genes estructurales de la vía de los tococromanoles, entre ellos *VTE3*, (Georgiadou y col. 2015, 2016). Adicionalmente, la baja heredabilidad del carácter podría ser debida a la existencia de interacciones genéticas (epistasis) entre distintos *loci* involucrados en la regulación de los contenidos de tocoferoles.

Varios reportes remarcaron el rol clave del gen *VTE3(1)* en la biosíntesis de tocoferoles, incluyendo el tomate. El estudio de los polimorfismos para este gen dentro de una población de 133 accesiones mediante el mapeo de QTL en semillas de *Brassica napus* demostró que más del 16.93% de la varianza asociada a la composición de tocoferoles y hasta un 10.48% del tocoferol total era explicado por polimorfismos en *VTE3* (Fritsche y col. 2012). Más recientemente en maíz, estudios de GWAS y mapeo de QTL lo identificaron también como un QTL mayor (Hong y col. 2018, Diepenbrock y col. 2017, Chander y col. 2008). Otro ejemplo de esto es el trabajo de Tang y col. (2016) en plantas transgénicas de *Lactuca sativa* para *VTE3* el cual contribuyó a incrementos del tocoferol total en hojas entre 50 a 84% como consecuencia de aumentos entre 60%-93% de γ -tocóferol y entre 41%-93% de α -tocóferol produciendo una mayor tolerancia al estrés lumínico. Sin embargo estos aumentos en tocoferoles repercutieron en una reducción en los contenidos de filoquinona (vitamina K) sugiriendo competencia por el sustrato común fitil difosfato, aunque los contenidos de vitamina A, carotenoides y clorofilas dependientes de este intermediario no presentaron diferencias. Otro dato remarcable es que la evaluación de la expresión de algunos genes estructurales de la vía de tocoferoles (*HPPD*, *HPT*, *TC* y γ -*TMT*) en estas plantas transgénicas de lechuga no presentaron diferencias con respecto a lo observado en los controles sin transformar.

Sumado a estos reportes en otras especies que muestran la importancia del rol del gen *VTE3* en la biosíntesis de tococromanoles, el trabajo de Almeida y col. (2011) lo reporta por primera vez en tomate junto con otros 16 QTL. Más recientemente, partiendo del mapeo fino de un QTL identificado en la línea de introgresión IL9.2.6, presentando un incremento del 70% en los contenidos de α - tocoferol y tocoferol total en frutos maduros de tomate, Quadrana y col.

(2014) describieron parte del mecanismo de regulación *VTE3(1)*, el cual puede explicarse por el nivel de metilación de su región promotora asociado a un elemento transponible del tipo SINE que además se regula diferencialmente en respuesta al ambiente. Estos resultados previos condujeron al estudio de su gen parálogo *VTE3(2)*, cuya expresión correlaciona con la acumulación de α - y tocoferol total como así también con otros genes que codifican para enzimas involucradas en la biosíntesis de estos compuestos en diferentes tejidos de tomate fuertemente (Quadrana et al 2013), apuntando a la expresión de *VTE3(2)* como un nodo de la regulación de este metabolismo. Además, la línea de introgresión (IL3.1) que porta un alelo silvestre del gen *VTE3(2)* presentó otro QTL para tocoferoles, lo cual reforzó su candidatura y llevaron a la formulación de las hipótesis que desarrollamos en este capítulo III. Principalmente, nuestro mayor interés fue evaluar la interacción entre estos dos genes y entre dos alelos.

El resultado inicial clave en el cual se apoyó la hipótesis de este capítulo fue la identificación de la presencia de un elemento transponible *SINE* en la región promotora de *VTE3(2)* similar al encontrado en *VTE3(1)* a partir de la exploración de su estructura génica. Además la identidad de secuencia entre estos parálogos es muy alta (con ausencia de polimorfismos que afecten la funcionalidad de la proteína), y mayor aun entre alelos, sugiriendo que las diferencias en la regulación transcripcional de ambos genes parálogos podría involucrar mecanismos epigenéticos.

Para poner a prueba esta hipótesis se diseñó y construyó una población segregante F2 a partir de la autofecundación de plantas (parcialmente) híbridas obtenidas al cruzar las IL3.1 y IL9.2.6.1. Mediante el genotipado de las mismas usando marcadores moleculares alelo-específicos para ambos parálogos diseñados *ad hoc*, pudimos diferenciar los alelos silvestres y domesticados de los genes *VTE3s*. El análisis de la población confirmó su segregación de acuerdo a las frecuencias esperadas. El análisis de los contenidos de tocoferol, como así también de la expresión y metilación de los genes *VTE3s*, en las plantas parentales en comparación con las líneas recombinantes F2 conteniendo los mismos genotipos sugirió interacciones entre el genotipo y el ambiente como así también una importante asociación entre la expresión de los genes *VTE3s* y los niveles de metilación de sus secuencias promotoras (figuras 26 y 27). Adicionalmente, la caracterización molecular de líneas de la población F2 con genotipos parentales reconstituyentes sugirió la existencia de interacciones entre ambos parálogos (Figuras 32 y 33)

Los resultados descriptos precedentemente sugieren que la inestabilidad a nivel transcripcional, y en menor medida en la relación $(\alpha+\gamma/\beta+\delta)$ de tocoferoles, podría ser debida a factores epigenéticos en trans establecidos en respuesta al cruzamiento entre los parentales IL3.1 y IL9.2.6.1 y/o durante las

meiosis entre los cromosomas homólogos en estado homocigota primero (líneas parentales) y heterocigotas luego (F1) que originaron la población F2 analizada en esta tesis. Alternativamente, la variación entre los genotipos parentales y sus respectivos genotipos reconstituidos en la población F2 podría deberse simplemente a variaciones intergeneracionales o la presencia de introgresiones de origen desconocido en los parentales que segregan independientemente de *VTE3(1)* y *VTE3(2)*.

Paso siguiente extendimos nuestro análisis a la población F2 completa (Figura 30), para poder cuantificar y evaluar no solo la contribución de cada factor a la variabilidad de la población, sino también de la interacción entre estos factores. Los modelos lineales mixtos (MLM) ya han sido propuestos como los más indicados para este tipo de análisis (Moore y col. 2018, Douglas y col. 2015). La exploración de la varianza fenotípica, explicada en gran medida por el ambiente y los niveles de metilación y expresión de los genes *VTE3s*, permitió entender la variabilidad de los datos fenotípicos.

Finalmente la disección e interrelación entre estas tres variables también mediante modelos MLM para los genes *VTE3(1)* y *VTE3(2)* expuso que la combinación de alelos *VTE3s* definida aquí como genotipo sería la variable de mayor importancia con consecuencias directas sobre la expresión y la metilación tanto en hojas como en frutos (Figuras 37). Este efecto de la combinación de alelos sobre la expresión al menos en hojas se puede observar en la figura 38, a través de la evaluación de correlación de Spearman entre todas las variables del modelo. No obstante, las correlaciones con los contenidos de tocoferol fueron bajas, esto posibilita pensar en el rol de otros genes involucrados además de los aquí estudiados. Por otro lado el efecto del ambiente sobre la expresión ocurriría indirectamente a través de la metilación del ADN (Figura 37). Es decir, el ambiente tiene un fuerte impacto sobre los niveles de metilación del ADN que a su vez regula la expresión de los genes. Debemos reconocer que nuestros análisis incluyeron tres niveles: metilación, expresión y la evaluación del fenotipo con lo cual no se descartan regulaciones a otros niveles intermedios aquí no explorados.

En conclusión, los resultados de este capítulo soportan la existencia de interacciones génicas entre los genes parálogos *VTE3(1)* y *VTE3(2)*. Estas se evidencian por la coordinación del nivel de expresión de los alelos de los genes *VTE3s*. Ambos parálogos se expresan diferencialmente entre tejidos y proponemos que cada uno evolucionó distintos mecanismo de regulación epigenética que permiten la modulación transcripcional de *VTE3(2)* en hojas y *VTE3(1)* en frutos. No obstante, sus consecuencias sobre los niveles de tocoferol requieren extender los análisis hacia otros otros actores además de los aquí estudiados.

CAPITULO IV - Integración del conocimiento - Perspectivas

El objetivo principal de esta Tesis fue aportar conocimiento para entender la regulación de la biosíntesis de tococromanoles, moléculas de suma importancia económica y nutricional dada la actividad de vitamina (E) que tienen en células de mamíferos. Desde su descubrimiento en 1922 (Evans y Bishop 1922), estas moléculas han sido solamente encontradas en organismos fotosintéticos y desde hace no mucho tiempo constituyen un blanco de manipulación tanto en alimentos como en productos comestibles. Para ello se escogió un sistema vegetal modelo, la planta de tomate, cuya importancia no es solamente económica sino también de creciente interés en la comunidad científica considerando la existencia de herramientas disponibles para la experimentación que existe para la especie. Con base en resultados previos del laboratorio (Almeida y col, 2011; Quadrana y col, 2012; 2014) se diseñaron experimentos que permitieran evaluar el efecto ambiental sobre la acumulación de estas moléculas en los frutos de tomate y la variación de las mismas en tejidos fotosintéticos.

Para esto se plantearon dos objetivos complementarios, los cuales fueron desarrollados en los capítulos II y III. El eje del capítulo II se focalizó en la búsqueda de nuevos determinantes genéticos involucrados en la biosíntesis de tococromanoles explorando nuevas variantes alélicas, más allá de las investigadas hasta la fecha proveniente de la especie silvestre *S. pennellii*, las cuales han explicado en parte, mucho del conocimiento obtenido hasta el momento sobre la regulación de la síntesis de tococromanoles en los frutos y hojas de tomate. En este capítulo se identificaron 60 nuevos *loci* junto con otros 8 QTL a partir de la aplicación de mapeo de asociación y mapeo por intervalos. Además de estos genes estructurales, se destacan 10 FT como nuevos candidatos que podrían estar regulando a aquellos primeros, un aspecto este mucho menos estudiado para esta y otras vías metabólicas de las que derivan compuestos de interés. Algunos de los *loci* cuyos productos corresponden a enzimas de las vías de tococromanoles y otras convergentes han sido ya reportados en maíz (Diepenbrock y col. 2017, Lipka y col. 2013, Hong Wang y col. 2018). Esto sugiere de alguna manera un alto grado de conservación de la vía, facilitando el entendimiento de los mecanismos regulatorios que están lejos de ser entendidos aún. Además de la exploración a nivel genómico y la arquitectura de la varianza fenotípica, el rol de la interacción genotipo ambiente y la estimación de la heredabilidad fueron estudiados en esta tesis. La cual se presentó con valores bajos para el contenido de tocoferol total y sus isoformas acentuándose drásticamente en plantas cultivadas en condiciones estresantes comparadas con el control. Esto constituye un resultado clave a la hora de abordar estrategias de mejoramiento con el objetivo de incrementar los contenidos de estos compuestos en tomate.

Esto dio como consecuencia el desarrollo del capítulo III. En el cual se presentan resultados de experimentos diseñados para intentar explicar las posibles causas de esta baja heredabilidad y encontrar nuevos elementos y herramientas a aplicar en programas de mejoramiento. Nos basamos en el análisis de los genes *VTEs* paralogos protagonistas en la vía. Estos resultados permitieron determinar una especificidad a nivel tejido en cuanto a los niveles de expresión; en frutos *VTE3(1)* parece tener un rol más relevante y en hojas es el parágo *VTE3(2)* que, de acuerdo a sus nivles de ARNm, sugieren un rol prioritario. Además el estudio conjunto de los niveles de metilación de sus regiones promotoras, expresión y evaluación del fenotipo, evidencio interacciones epistáticas entre ambos regulada por mecanismos epigenéticos y determinada a través de varios son los factores: i) combinación de alelos *VTE3s* ii) el manejo y iii) el ambiente. Estos experimentos revelaron que la modulación transcripcional de *VTE3(1)* y *VTE3(2)*, mediada en parte por la variación en los niveles de metilación de sus promotores, es en efecto, un blanco importante para el mejoramiento de este carácter nutricional.

En conjunto, esta tesis pone en evidencia la importancia de explorar tanto la variación intraespecífica como interespecífica para caracterizar las bases genéticas de fenotipos metabólicos y la necesidad de incluir la variación epigenética como un factor determinante de caracteres de interés agronómico. La comprensión de estos mecanismos permitirá diseñar mejores estrategias a ser utilizadas en programas de mejoramiento tendientes a seleccionar cultivares de tomate (u otras especies) con mayores contenidos de tococromanos.

CAPITULO V – Materiales y Métodos

Este capítulo resume los materiales y métodos utilizados con el fin de abordar las hipótesis propuestas en esta tesis (ver Tabla S1 en el Anexo sección IV).

V.1 Descripción de los materiales y diseño experimental de una población de GWAS y 124 líneas recombinantes derivadas de un cruzamiento multiparental cultivados en ambientes contrastantes para la búsqueda de QTLs

V.1 .1 Materiales GWAS

La colección completa de genotipos de la población de GWAS consiste en 210 entradas en total, de los cuales 146 son utilizados en esta Tesis (Tabla S2, suplementaria). Entre estas, 10 accesiones pertenecen a *S. pimpinellifolium* (SP), entre las cuales 6 son originarias de Perú y 2 de Ecuador y 1 de México. Un total de 107 accesiones son *Solanum lycopersicum* var. *Cerasiforme* (SLC). Entre ellas, 15 Originarias de Perú, 26 de México, 2 de Colombia, 3 Bolivia, 3 Brasil, 1 Argentina, 1 Belice, 1 Dinamarca, 2 Costa Rica, 2 de Cuba, 1 de Checoslovaquia, 10 de Ecuador, 1 Alemania, 1 Indonesia, 81 de áreas subtropicales. Finalmente, 20 accesiones pertenecen a un grupo genético mixto donde están incluidos materiales comerciales de tomates cherry de la colección de INRA que incluyen materiales *S. pimpinellifolium*, *Solanum lycopersicum* var. *cerasiforme* and *S. lycopersicum* var. *lycopersicum*. Algunas accesiones son originarias de climas secos como las costas secas de Perú mientras que otras de los valles húmedos de Ecuador. Estas accesiones fueron particularmente incluidas en el panel con el interés de aumentar la diversidad genética frente déficit hídrico.

La colección de semillas fueron provistas por distintos bancos de germoplasma: Centro INRA avignon (CRB-Leg, Francia), Centro de la Universidad de Davis (TGRC, USA), Departamento de Biología Molecular y Bioquímica de la Universidad de Málaga (España), el Instituto de investigación de la Industria de las plantas de Vavilov San Petersburgo (VIR, Rusia), El instituto de conservación, Mejoramiento y Agrobiodiversidad de Valencia (COMAV, España) y el centro de Reservas genéticas de Wageningen (CGN, Holanda).

V.1 .2 Materiales MAGIC POP

La población multiparental está compuesta por 124 F7 líneas recombinantes derivadas de cruces de 8 líneas parentales escogidas de un conjunto de 360 accesiones (Rancet y col. 2008) con el propósito de maximizar la diversidad genética. Para su construcción se realizaron 4 cruces entre *S. lycopersicum* var. *esculetum* y *S. lycopersicum* var. *cerasiforme*, con frutos de menor y mayor tamaño respectivamente (Figura 14, capítulo II). Los cuatro genotipos del grupo de *S. lycopersicum* elegidos fueron: Levovil (L1), Stupicke (L2), LA0147 (L3) y

Ferum (3). Del grupo de *S. lycopersicum* var. *cerasiforme*, se escogieron los genotipos Cervil (C1, Criollo (C2), Plovdiv (C3) y LA1420 (C4). Posteriormente, los híbridos F1 Levovil x Cervil y Stupicke PR x Criollo fueron cruzados para obtener 120 plantas provenientes del doble cruce de híbridos F1. En paralelo, las líneas híbridas LA0147 x Plovdiv24A y Ferum x LA1420, también fueron cruzadas para obtener otro grupo de 120 plantas doble híbridas F1. Finalmente, estos dos grupos de 120 plantas doble híbridas fueron cruzadas independientemente. Dos descendientes por cruce fueron obtenidas produciendo 480 plantas F1, las cuales poseen regiones para cada uno de los 8 parentales fundadores en sus genomas. Estas plantas fueron propagadas a través de semillas durante 6 generaciones creando un conjunto de 397 líneas MAGIC F7, de las cuales 124 fueron analizadas para el mapeo de QTL asociadas a contenidos de tococromanos de esta tesis. Las mismas fueron utilizadas para el mapeo de QTL en análisis previos identificando cientos de QTLs para caracteres sensoriales, físicos y químicos de frutos de tomate medidos en condiciones óptimas de riego (Causse y col. 2001, Saliba-Colombani y col. 2001, Lecomte y col. 2004).

V.1 .3 Diseño experimental y condiciones de riego, ensayos GWAS y MAGIC-POP

La población de GWAS y las 124 líneas F7 fueron crecidas en invernadero en Aviñón (Francia) de marzo a julio del 2015 en condiciones de control (figura 39). En el caso de los 146 genotipos del GWAS además se realizó un tratamiento de sequía. El tratamiento control fue seteado de acuerdo a datos de evotranspiración y coeficientes de medios de cultivo de tomate para condiciones de invernadero (FAO Water, 2015). El drenaje máximo de 25% y una humedad del sustrato de 65% para el tratamiento control. El tratamiento sequia fue progresivamente aplicado luego de floración del segundo racimo tomando como punto de inicio aquellos genotipos de temprana floración. Se comenzó con una reducción del riego de 25% en comparación al control durante la primera semana y 40% durante el resto del experimento. Permitiendo generar un moderado déficit hídrico. La intensidad de luz fue de 700 W/m² y reducida durante los picos de calor durante el verano. Las plantas fueron crecidas en macetas plásticas de 4 litros con turba como sustrato y una solución nutritiva con la siguiente composición (2, 4, 6 mmol l⁻¹, N, P, y K, respectivamente) fue administrada. La humedad relativa de la turba fue monitoreada utilizando como medidor el GRODAN®. Tres plantas por tratamiento por accesión fueron aleatorizadas en el invernadero en el caso del GWAS y dos plantas en el caso de las 124 F7 MAGIC-POP.



Figura 39. Fotografía del ensayo conducido en INRA Aviñón en junio 2015. La fila de plantas derecha e izquierda corresponden a los mismos genotipos en condiciones control y sequía (-40% de agua) respectivamente (adaptado de Albert y col. 2017).

V.1.4 Mapeo y datos genotípicos GWAS

La población de GWAS fue genotipada a partir del Tomato Infinium Array desarrollado dentro del proyecto SolCAP project (<http://solcap.msu.edu/>). Este consta de 8,784 SNP sobre la totalidad del genoma de tomate, entre estos 7663 SNP pasaron los test de calidad (Hamilton y col. 2012; Sim y col. 2012). Luego del filtrado por datos faltantes y menor frecuencia alélica (MAF), la matriz final consta de 6100 SNP. La línea de corte utilizada para el filtrado por menor frecuencia alélica (MAF) fue de 0,04 de acuerdo a Aulchenko y col. (2007). Previo a cualquier análisis genético los datos de SNPs faltantes fueron reemplazados por aquellos alelos de mayor frecuencia alélica. Los SNPs fueron renombrados de acuerdo a las posiciones en la versión SL2.50 del genoma de tomate como S01_58000085 para el para la posición 58000085 pares de bases sobre el cromosoma 1.

V.1.5 Mapeo y datos genotípicos MAGIC POP

Para la selección de los 1.345 utilizados en esta tesis para la determinación de QTLs asociados a los contenidos de tococromanoles de las 124 líneas utilizadas, se partió de un total de más de 4 millones de SNPs detectados en el trabajo de Causse y col. (2013) el cual consistió en la resecuenciación de las 8 líneas (los datos crudos se encuentran depositados en ENA con el número de

accesiones ERR327646 hasta ERR327656, junto a los datos de SNPs e *INDELs* numero de accesiones, ERZ015686 a ERZ015701). Tres pasos de filtrado fueron realizados. El primero basado en los SNPs de mejor calidad en cuanto a la predicción de su reproducibilidad, el segundo paso consistió en filtrado por posición de cada SNP manteniendo solo un SNP por set de datos para el mismo perfil alélico y finalmente en el tercer paso fueron seleccionados en base a intervalos con una distancia máxima de 8 cM. La distancia genética entre marcadores fue calculada para cada posición física del genoma por regresión lineal utilizando solo los 5 marcadores mas cercanos del EXPEN 2000map (solgenomics.net).

V.1.6 Identificación de asociaciones por GWAS a través de modelos MLMM (Multitrait Linear Mixed Model); cálculos de heredabilidad y porcentaje de la varianza explicada

Previo a la realización de los ANOVA a partir de los datos fenotípicos crudos de los 146 materiales de GWAS para contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y en frutos maduros se realizó la normalización aplicando la transformación de Box Cox (1964) ya que los 6 caracteres evaluados presentaron distribuciones distintas a la normalidad y no homocedaceas sus varianzas. El cálculo del ANOVA se realizó sobre el modelo siguiente $P_{im} = G + M + E + G \times E + R$; donde P indica el valor fenotípico medio para la accesión i en la condición m, G el genotipo para esa accesión, M el efecto de lugar de medición, GxE la interacción genotipo ambiente y R la variación residual o error del modelo. Los modelos tuvieron en cuenta la variable operacional M, como factor de ajuste, separando de esta manera variación debida a efectos técnicos de medición a excepción del GWAS 2. Para este ajuste de los datos se utilizó el paquete de R lme4 (Douglas y col. 2015) específico para la aplicación de modelos lineales mixtos. Los distintos GWAS fueron llevados a cabo con corrección de la estructura Q y modelado de la covarianza con la matriz de parentesco K estas matrices fueron implementadas modificando el modelo MLMM descrito por Segura y col. (2012) que tiene en cuenta la estructura de las poblaciones como cofactor el script de R con la función mlmm_cof.r se puede obtener en (<https://cynin.gmi.oeaw.ac.at/home/resources/mlmm>). Ambas matrices fueron calculadas de la misma forma que en Albert y col. 2016. Brevemente el modelo MLMM está basado en la librería Emma (Kang y col. 2008) el mismo tiene en cuenta múltiples modelos mixtos de regresiones lineales por pasos, con inclusión y exclusión de cofactores reestimando la componente de la varianza en cada paso. Este modelo incrementa el poder de detección y reduce el FDR cuando lo comparamos con los modelos simples tradicionales. Dos criterios de selección del modelo fueron implementados para la evaluación de los modelos test de BIC y extBIC, criterios de información bayesiano (Schwarz y Gideon 1978). El porcentaje de la varianza explicada fue

obtenido para cada carácter fue calculado en el paso optimo del modelo MLM y La heredabilidad de cada carácter fue computada en el paso 0 del modelo MLM donde no incluye aun ningún marcador al modelo MLM. Los 4 análisis de identificación de asociaciones para cada uno de los caracteres en ambas condiciones, control y estrés e independientemente del ambiente indicada como “ambos” fue llevado a cabo con distintos grados de astringencia. El seteo de estos 4 análisis de GWAS vario en 2 variables: i) número de cofactores del modelo los cuales explican la varianza fenotípica asociada, siendo 5 cofactores para los análisis de GWAS 1, 2 y 3 y 10 cofactores, en el caso del análisis de GWAS 4 ii) la otra variable fue el p-valor de significancia de corte para determinar que una asociación sea verdadera, la misma vario entre 10^{-5} para el análisis de GWAS 1 y 10^{-4} para los tres restantes análisis de GWAS 2,3,4.GWAS.

V.1.7 Determinación de los IC (intervalos de confianza) que definen los QTL del GWAS

Para definir los intervalos de confianza que determinan los QTLs del análisis de GWAS estos fueron basados a partir del LD entre pares de marcadores considerando los 6100 SNPs de la matriz de datos genómicos, tal como se calculó en Albert y col. (2017) implementando el paquete *genetics* (Warnes y Leisch 2012). El cálculo para el estimador de r^2 estimado critico se realizó en base a 100000 marcadores aleatorios al azar de *loci* no ligados en diferentes cromosomas. El percentil 95 de la distribución de r^2 no ligados fue 0.28, este valor fue considerado como LD crítico, para definir si dos marcadores están o no ligados. Luego para cada uno de los marcadores que presentaron asociaciones significativas, se recorrió en sentido rio arriba y abajo de los cromosomas definiendo de esta forma los intervalos tomando el último marcador ligado respecto del SNP significativo. El número de genes dentro de cada intervalo fue definido en la versión del genoma del tomate (ITAG2.4).

V.1 .8 Determinación QTLs de la F7 MAGICPOP

Para la detección e identificación de los polimorfismos casuales de las variaciones fenotípicas se realizó, lo que se denomina un mapeo de intervalo simple utilizando la herramienta mpIM de R (Huang y George 2011). Esta función, computa el efecto de cada parental en base a una regresión de múltiples probabilidades, función mpprob, con ventanas de un tamaño de 2 centimorgans, las cuales definen la resolución del mapeo. Para evaluar el grado de significancia del efecto parental para cada uno de los QTLs identificados se utilizó el Test de join Wald con un p valor de 1.72×10^{-4} . Se utilizó la función sim.sigthr luego de computar 1000 permutaciones en todo el genoma al azar determinando un p-valor de corte 0,05.

V.1.9 Búsqueda de Factores de Transcripción

Con el objetivo de identificar los locus correspondientes factores de transcripción identificados a través de las dos metodologías de mapeo, GWAS2 y mapeo de QTLs de la MAGICPOP F7. Se realizó una búsqueda en dos pasos, primero se identificaron TFs sobrerrepresentados. Para esto se partió del mismo listado de 86 genes relacionados al metabolismo de tococromanos utilizado para evaluar el enriquecimiento de genes VTE (Tabla S5 suplementaria) el cual fue basado en genes VTE de la literatura utilizando la herramienta provista por base de datos la PlantTFDJB (Jin y col. 2017). Esta herramienta de identificación de enriquecimiento, permite encontrar TFs sobrerrepresentados de unión a motivos CIS de las región promotora para conjuntos de genes targets a partir de la base de datos de regulaciones transcripciones reportadas en la literatura, ensayos de ChiP-seq y otros experimentos que permiten identificar la unión de TF a sitios específicos del DNA. Para verificar el grado de significancia de enriquecimiento de TFs, se utilizó la prueba exacta de Fisher, contrastando contra todos los motivos de la base de datos de regulaciones previamente reportadas.

V.2.1 Descripción de la obtención de una población F2 proveniente de la cruce de dos líneas de introgresión

Las semillas de las M82 (accesión LA3475) y las líneas IL3.1 (accesión LA4043), IL9.2.6 (accesión LA4083) para la construcción de la población de 80 plantas recombinantes F2 a partir de la cruce de las líneas introgresion IL3.1 y IL9.2.6 fueron obtenidas del Centro de recursos genéticos de la Universidad de Davis (<http://tgrc.ucdavis.edu>). Estas líneas de introgresion fueron desarrolladas por Eshed y Zamir, 1995. La línea IL9.2.6.1 fue obtenida a partir de una población segregante derivada del cruzamiento de la IL9.2.6 (accesión LA4083) con el parental M82 (accesión LA3475) (Quadrona y col. 2014). Para la obtención de la F1 se cruzó la línea IL3.1 (accesión LA4043) con la IL9.2.6.1 (ver cruzamientos debajo). La cruce F1 dio origen a la población recombinantes F2 cruce IL3.1 x IL9.2.6.1 de la cual se genotiparon 80 plantas para el ensayo denominado “apilamiento de dos loci” (Figura 30. Esquema de obtención de los 9 genotipos recombinantes, sección III.2.3.).

V.2.2 Cruzamientos

Los cruzamientos se realizaron mediante emasculación (escisión de los órganos reproductivos masculinos) de la flor aceptora de polen previa a la apertura de la flor. Uno o dos días posteriores se tomó de la planta dadora de polén, una flor totalmente desarrollada y un día antes de la anthesis. Con el uso de una pinza de punta fina, se retiró el cono formado por las anteras (procurando que el pistilo no quede dentro del cono) y se depositó en la flor

previamente emasculada. De esta manera nos aseguramos que i) la flor quede cubierta evitando que otros granos de polen se depositen en el pistilo y ii) que este, una vez maduro y receptivo, se encuentre colmado de granos de polen de la planta deseada. Las flores cruzadas fueron identificadas con una marca en el pedúnculo y una tarjeta indicando la fecha y procedencia del polen.

V.2.3 Desarrollo y aplicación de marcadores moleculares para el genotipado de la población recombinante F2 Cruza IL 3-1 x IL9.2.6.1

Se obtuvo una población de 80 líneas recombinantes a partir de un cruzamiento entre la IL9.2.6.1 (que porta el alelo silvestre del gen *VTE3(1)*) con la IL3.1 (que porta el alelo silvestre de su parólogo *VTE3(2)*). Esta población fue cultivada en condiciones controladas de invernadero durante el verano (2013-2014) y en el estadio de plántula se realizó el genotipado de cada individuo conjuntamente con 10 individuos de los progenitores (IL9.2.6.1 y IL3.1) y del parental recurrente *S. lycopersicum* cv M82. Para esto se diseñaron dos marcadores de PCR a partir de las secuencias genómicas de *VTE3(1)* y *VTE3(2)* permitiendo discriminar los 2 alelos posibles (*S. lycopersicum* -L- y *S. pennellii* -P-) para cada uno de los 2 parálogos por tamaño, estos *primers* se encuentran descritos en la Tabla S7 listado de *primers*. El análisis se realizó sobre ADN genómico de los individuos, los que previamente fueron seleccionados por su calidad y cantidad. A su vez, un 17,5% de los individuos (14 individuos) fueron genotipados por triplicado y el 100% por duplicado. La segregación encontrada en los 80 individuos recombinantes resultó de acuerdo a lo esperado (Chi cuadrado $X^2= 6,35$ siendo el valor umbral de 20,1 $p<0,001$ (Tabla 12).

Para analizar ambos marcadores se realizó una PCR a partir de 100ng de DNA genómico de hojas de plántulas (extraído con el kit Phytopure de GE Health Care) con TAQ polimerasa (Invitrogen) y *primers* específicos (ver Tabla S7 Anexo IV) en un volumen final de 20 μ l, con el siguiente programa de ciclado: 5' 95°C y 40 ciclos de 30'' 95°C, 30'' 59°C, 2' 72°C. Los productos de PCR fueron analizados por electroforesis en gel de agarosa (1 %) teñido con bromuro de etidio.

Tabla 12. Genotipado de la población cruza IL3.1 x IL9.2.6.1 para los genes *VTE3(1y2)*

Genotipo ^(a)	Combinación de alelos ^(b)	FE ^(c)	NPE ^(d)	N° plantas observadas ^(e)	$\chi^2 = (O-E)^2/E^{(f)}$
R1	P ₁ P ₁ L ₂ L ₂	1/16	5	8	1,8
R2	P ₁ P ₁ P ₂ L ₂	2/16	10	7	0,9
R3	P ₁ L ₁ L ₂ L ₂	2/16	10	14	1,6
R4	P ₁ L ₁ P ₂ L ₂	4/16	20	19	0,05
R5	P ₁ P ₁ P ₂ P ₂	1/16	5	6	0,2
R6	P ₁ L ₁ P ₂ P ₂	2/16	10	11	0,1
R7	L ₁ L ₁ L ₂ L ₂	1/16	5	5	0
R8	L ₁ L ₁ P ₂ L ₂	2/16	10	7	0,9
R9	L ₁ L ₁ P ₂ P ₂	1/16	5	3	0,8
Total	-	1	80	80	6,35

^(a)Genotipo de las plantas F2 recombinantes cruza (IL9.2.6.1x IL3.1) según los marcadores ligados físicamente a *VTE3(1)* y *VTE3(2)*. ^(b)Combinación alélica para los genes parálogos *VTE3(1)* y *VTE3(2)* respectivamente e indicado con el subíndice. P corresponde *Solanum Pennelli* y L corresponde *Solanum lycopersicum*. ^(c)Frecuencia esperada en el genotipado para una segregación mendeliana. ^(d)Número de plantas esperadas en el ensayo para una segregación mendeliana. ^(e)Número de plantas observadas. ^(f) Prueba de χ^2 para 80 plántulas F2, El valor umbral de χ^2 para 8 grados de libertad (n-1) con un p-valor de 0,01 es 20,1. Los datos se ajustan a la distribución teórica esperada, o lo que es lo mismo, no hubo distorsión en la segregación.

V.2.4 Diseño experimental y condiciones de la población recombinante F2 Cruza IL 3-1 x IL9.2.6.1

Las 80 plantas recombinantes Cruza IL 3-1 x IL9.2.6.1 se cultivaron en macetas de 20 litros bajo condiciones controladas en invernáculo con un fotoperíodo de 16/8 h, temperatura: 25°C ± 5°C, humedad: 60% e irradiancia: 300 ± 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Para todos los casos, los muestreos se realizaron al medio día para asegurar el máximo nivel de irradiancia. Solo los pericarpios de frutos (sin piel, placenta ni semillas) recién cosechados fueron muestreados e inmediatamente congelados en N2 líquido y de igual forma para las hojas. Las muestras fueron conservadas a -80°C hasta su uso. Para el experimento contraste en condiciones de campo, se partió de clones (esquejes) de las mismas 80 plantas recombinantes Cruza IL 3-1 x IL9.2.6.1, se crecieron en macetas de 150 ml hasta alcanzar raíces y se trasplantaron a tierra bajo condiciones de producción a campo, en el Campo Experimental del Instituto de Horticultura de la Universidad Nacional de Cuyo, Mendoza, Argentina, 32°50'S, 68°52'W y a 900 msnm. El experimento de campo se realizó entre octubre de 2015 y marzo de 2016 en un diseño aleatorio de tres repeticiones con tres plantas cada una. El experimento fue protegido con

una malla antigranizo y el riego se aplicó de tal manera de mantener constante el contenido de agua disponible en el suelo. Los frutos maduros por planta se cosecharon aproximadamente 60-65 días después de la antesis.

V.3 Medición de datos fenotípicos (contenido de tococromanoles) por HPLC.

La medición del contenido de tocoferoles y tocotrienoles fue realizada para las todas las muestras fue realizada preservado inmediatamente en N₂ líquido, molido y conservado a - 80°C hasta su posterior liofilización.

La extracción de tococromanoles se realizó como se describe en Fraser y col. (2000) con las siguientes modificaciones: 500 mg de fruto o 100 mg de hojas se extrajeron con 1,5 ml de metanol, se vortexeo vigorosamente la mezcla, y se añadió 1 ml cloroformo. Después de 5 minutos de sonicación, se agregó 1 ml de buffer Tris (pH 7,5 50 mM / NaCl 1 M). La fase clorofórmica se recuperó y la fase metanólica (conteniendo el sedimento remanente) se extrajo de nuevo con 2 ml de cloroformo. Ambos extractos de cloroformo se combinaron y se ajustó a un volumen final de 4 ml. Se tomaron 3ml de extracto y se secaron bajo nitrógeno gaseoso. Momentos antes de inyectar en el HPLC, se tomaron los extractos secos y se resuspendieron en 0,2 ml de fase móvil.

El contenido de tocoferoles se determinó usando un HPLC Hewlett-Packard serie 1100 acoplado con un detector de fluorescencia (Agilent Technologies serie 1200). La separación se llevó a cabo en una columna LiChrospher® 100 DIOL (5 µm) LiChroCART® 250-4 (Agilent Technologies), mantenida a temperatura ambiente usando un sistema de solvente isocrático (fase móvil) que consta de 99,5:0,5 heptano/ Ter-butil- metil-eter con una velocidad de flujo de 1,6 ml / min. Los compuestos de elución fueron detectados y cuantificados por fluorescencia con excitación a 296 nm y emisión a 340 nm. La identificación y cuantificación de los diferentes tocoferoles se logró mediante la comparación con los tiempos de retención y las áreas de los picos de los estándares adquiridos en Merck (Calbiochem # 613424). Una curva de calibración diaria se llevó a cabo utilizando una solución mezcla de tocoferoles (α , β γ y δ) y tocotrienoles (α y δ) con un rango de concentración entre 0,31 y 5 mg / ml para cada isoforma.

V.4.1 PCR en tiempo Real

Los niveles de ARNm se cuantificaron mediante PCR cuantitativa en tiempo real usando un termociclador en tiempo real (Applied Biosystem 7500). Para la reacción de PCR se utilizó el kit SYBR (Applied Biosystem), el cual contiene una TAQ polimerasa con propiedad HOT-START (la cual sólo se activa tras

una incubación a 95°C), Mg²⁺, dNTPs y buffer. Los *primers* específicos se utilizaron a concentraciones variables dependiendo del nivel de expresión del gen a analizar cómo se detalla en la ver Tabla S7 Anexo IV. El programa de ciclado para la RT-qPCR fue el siguiente, 1 ciclo de 95°C, 5 minutos. 40 ciclos de 95°C, 15s; Ta°C 1min, donde la temperatura de *annealing* (Ta) fue específica para cada gen, como se especifica en la Tabla S7 Anexo IV. Al finalizar el programa de ciclado se realizó una curva de disociación para corroborar la especificidad de la PCR. Para cada gen analizado los datos de fluorescencia en función del ciclo de amplificación de cada reacción se exportaron en un archivo .csv.

V.4.2 Cálculo de la expresión de los genes parálogos VTE3(1) y VTE3(2) en ambas condiciones (campo e invernáculo) y ambos tejidos (hoja y frutos) de la población F2 cruza IL3.1 X IL9.2.6.1 basada en la eficiencia de los primers alelo específicos

La expresión génica normalizada se calculó por el método de delta delta Ct tomando alguna de las mejoras propuestas por (Hellemans y col. 2007) y adaptada al caso en particular que requiere comparar la expresión de un alelo particular con respecto a los otros tres alelos restantes, la ecuación fue la siguiente:

$$\text{Log}_2 \left[\frac{E_{ijk}^{(\bar{C}t_{(4 \text{ alelos; muestras F2; tejido } k)} - \bar{C}t_{(\text{alelo } i; \text{muestra } j; \text{tejido } k)})}}{E_{rjk}^{(\bar{C}t_{(\text{gen } r; \text{muestras F2, tejido } K)} - \bar{C}t_{(\text{gen } r; \text{muestra } j; \text{tejido } k)})}} \right]$$

En esta se transformó, la expresión relativa en base al logaritmo de base 2. La expresión relativa esta dada por el delta delta ct de: la relación de uno de los cuatro alelos VTE3s denominado como i respecto de la media de los cuatro alelos en toda la F2 para una muestra dada j y tejido k (que puede ser hoja o fruto) relativizándolo a la eficiencia del gen de referencia r en la misma muestra j y tejido k con respecto a la media de los ct de toda la F2 del gen de referencia.

A continuación se detalla cada término de la ecuación:

E_{ijk}: es la eficiencia de amplificación del alelo target i, para la muestra recombinante j en un tejido k (que puede ser hoja o fruto).

E_{rjk}: es la eficiencia del gen de referencia r o housekeeping para la muestra recombinante j en el tejido k.

Ct (4 alelos; muestras F2; tejido k): valor promedio de los ct de los cuatro alelos de para toda la F2 en el tejido k, hoja o fruto.

Ct (alelo i; muestra j; tejido k): valor promedio de los ct de tres replicas biológicas para el alelo i de la muestra j en el tejido k, hoja o fruto.

Ct (gen r; muestras F2, tejido k): valor promedio de los ct de toda la F2 para el gen de referencia, muestra j en el tejido k, hoja o fruto.

Ct (gen r; muestra j; tejido k): valor promedio de los ct de tres replicas biológicas para el gen de referencia en la muestra j y tejido k, hoja o fruto.

Como gen de referencia o *housekeeping* en este caso se usó el factor de elongación E1 α el cual fue puesto a punto en (Quadrana y col. 2013). Las eficiencia para cada gen fue calculada utilizando el programa LinReg (Ruijter y col. 2009). Para el caso de las plantas heterocigotas F2 (R2, R3, R4, R6, R8) para poder realizar comparaciones entre alelos para un mismo gen. Se verificó que la eficiencia amplificación de los amplicones particular, de cada alelo *S. Pennelli* y *S. Lycopersicum*, sea igual a la suma de ellos.

Técnicamente se chequeo que el valor de expresión relativo al amplicón que no diferencia *VTE3(1)pen* de *VTE3(1)Lyc* e amplifica a ambas regiones, no presenté diferencias con respecto a la suma de los amplicones particulares. Con lo cual que para estas plantas (heterocigotas) se analizaron tres sets de *primers*, uno para *VTE3(1)pen* otro *VTE3(1)Lyc* y un último par que amplifica ambos amplicones simultáneamente *VTE3(1)pyL*. Esto mismo se hizo para el otro parólogo, *VTE3(2)*.

V.5 Análisis del nivel de metilación de la región promotora de los genes parálogos VTE3(1y2) mediante McrBC-PCRsc. Para la población F2: cruza IL3.1 X IL9.2.6.1, se evaluaron los 4 alelos, en dos condiciones (campo e invernáculo) y en dos tejidos (hoja fuente y frutos maduros)

A partir de 100 mg de tejido se realizó la extracción de ADN genómico con el kit PhytoPure (GE Healthcare). 1 μ g de ADN fue digerido con 10 unidades de enzima McrBC (NEB) por 3 hs. a 37°C, en paralelo se realizó el control de ADN sin digerir.

Para el análisis del ensayo con McrBC, se utilizó la reacción en cadena de la polimerasa (PCR) mediante la utilización de *primers* alelo específicos para *S. lycopersicum* y *S. pennelli* para las regiones *SINE* reportada (productos 450 pb) y *SINE-TSD* (productos 793pb y 934pb respectivamente) reportada en

<http://solgenomics.net> como DMR 09944 (los *primers* se encuentran listados dentro de la Tabla S7 sección anexo), 50 ng de ADN tratado con McrBC fue amplificado por 28 ciclos con 1X del buffer de reacción, 0.75mM MgCl₂, 0.1μM dNTPs, 0.4μM *primers* específicos y 1 unidad de Polimerasa ADN Taq Platinum (Invitrogen). El ciclo fue: 15'' 94°C, 1'55°C y 1' 72°C. En simultáneo fue realizada la PCR para el ADN no tratado con la enzima. Se utilizó como control de amplificación *primers* específicos de rubisco cloroplastídica (*rbcl*) para todos los ADN (tratados y no tratados). Como control de la eficiencia de corte de la enzima endonucleasa se trató un de plásmido linealizado. Los productos de PCR fueron corridos en un gel 1% de agarosa, las fotos fueron obtenidas con el GBOX y la cuantificación de la intensidad de banda fue realizada con el Image J. Los datos de metilación del promotor están expresados tomando como 0% de metilación basado en la intensidad de banda del control no tratado.

CAPÍTULO VI – Referencias bibliográficas

Abushita A, Hebshi E, Daood H, Biacs P: **Determination of antioxidant vitamins in tomatoes.** Food Chem 1997, 60, 207–212.

Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin G, Tanksley S: **Transcriptome and Selected Metabolite Analyses Reveal Multiple Points of Ethylene Control during Tomato Fruit Development.** The Plant Cell 2005, 17(11), 2954–2965. doi:10.1105/tpc.105.036053

Albert E, Gricourt J, Bertin N, Bonnefoi J, Pateyron S, Tamby J-P, Bitton F, Causse M: **Genotype by watering regime interaction in cultivated tomato: lessons from linkage mapping and gene expression.** Theoretical and Applied Genetics 2016; 129, 395–418.

Alexander L & Grierson D: Ethylene biosynthesis and action in tomato: **A model for climacteric fruit ripening.** Journal of Experimental Botany 2002, 53(377), 2039–2055. doi:10.1093/jxb/erf072

Alleman M, Doctor J: **Genomic imprinting in plants: observations and evolutionary implications.** Plant Mol Biol. 2000, 43(2-3):147-61.

Almeida J, Azevedo S, Spicher L, Glauser G, Von Dorp K, Guyer L, Del Valle Carranza A, Asis R, de Souza AP, Buckeridge M, Demarco D, Bres C, Rothan C, Pers LE, Hortensteiner S, Kessler F, Dormann P, Carrari F, Rossi M: **Down-regulation of tomato strongly impairs tocopherol biosynthesis and affects prenillipid metabolism in an organ-specific manner.** Journal of Experimental Botany 2015, 67(3):919-34. Doi 10.1093/jxb/erv504

Almeida J, Quadrana L, Asís R, Setta N, De Godoy F, Bermúdez L, Otaiza SN: **Genetic dissection of vitamin E biosynthesis in tomato.** Journal of experimental botany 2011; 1-18. doi:10.1093/jxb/err055

Almeida J, Da Silva A, Spicher L, Glauser G, Vom Dorp K, Guyer L del Valle Carranza, Asis R. Pereira de Souza A, Buckeridge M: **Down-regulation of tomato PHYTOL KINASE strongly impairs tocopherol biosynthesis and affects prenillipid metabolism in an organ-specific manner.** J. Exp. Bot. 2016, 67, 919–934

Alonso-Blanco C, El-Assal SE, Coupland G, Koornneef M: **Analysis of natural allelic variation at flowering time loci in the Landsberg erecta and Cape Verde Islands ecotypes of Arabidopsis thaliana.** Genetics 1998, 149(2):749 -764.

Alseekh S, Ofner I, Pleban T, Tripodi P, Di Dato F y col: **Resolution by recombination: breaking up Solanum pennellii introgressions.** Trends Plant Sci 2013. 18:536-38

Asimit J, Zeggini E: **Rare variant association analysis methods for complex traits.** Annu Rev Genet 2010, 44:293-308.

Atwell S, Huang YS, Vilhjalmsdottir BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT y col: **Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines.** Nature 2010, 465(7298):627-631

Aulchenko Y, Ripke S, Isaacs A, van Duijn CM: **GenABEL: An R library for genomewide association analysis**. Bioinformatics 2007, 23, 1294–1296.

Auldrige M, Block A, Vogel J, Dabney-Smith C y col: **Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family**. The Plant Journal 2006, 45(6), 982–993. doi:10.1111/j.1365-313X.2006.02666.x

Azzi A: **Molecular mechanism of alpha-tocopherol action**. Radic Biol Med. 2007, 1;43(1):16-21

Bai Y, Lindhout P: **Domestication and breeding of tomatoes: what have we gained and what can we gain in the future?** Ann Bot. 2007, 100(5):1085-94.

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO: **QTL mapping in new Arabidopsis thaliana advanced intercross - recombinant inbred lines**. PLoS One 2009, 4(2):e4318.

Ballester AR, Molthoff J, de Vos R, y col: **Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SIMYB12 leads to pink tomato fruit color**. PLANT PHYSIOLOGY 2010, 152, 71–84.

Barrantes W, Lopez Casado G, Garcia-Martinez S, Alonso A, Rubio F y col: **Exploring new alleles involved in tomato fruit quality in an introgression line library of Solanum pimpinellifolium**. Front. Plant Sci 2016, 7:1172

Barry CS, McQuinn RP, Chung MY, Besuden A, Giovannoni JJ: **Amino acid substitutions in homologs of the STAY-GREEN protein are responsible for the green-flesh and chlorophyll retainer mutations of tomato and pepper**. Plant Physiology 2008, 147, 179–187.

Barry CS, Aldridge GM, Herzog G, Ma Q, McQuinn RP, Hirschberg J, Giovannoni JJ: **Altered chloroplast development and delayed fruit ripening caused by mutations in a zinc metalloprotease at the lutescent2 locus of tomato**. Plant Physiology 2012, 159, 1086–1098.

Baulcombe DC, Dean C: **Epigenetic regulation in plant responses to the environment**. Plant Biotechnol J. 2011 Cold Spring Harb Perspect Biol. 2014,2;6(9):a019471 doi: 10.1101/cshperspect.a019471.

Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, Campos J, Araujo II, Anderson TM, Vilhjalmsen BJ y col: **Genetic architecture of skin and eye color in an african-European admixed population**. PLoS Genet 2013, 9(3):e1003372

Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier K, Groot SPC, Koornneef M: **Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis**. Plant Physiol 2000, 124:1595-

Bodmer W, Bonilla C: **Common and rare variants in multifactorial susceptibility to common diseases**. Nat Genet 2008, 40(6):695-701

Bolger T y col: **The genome of the highly stress tolerant wild tomato species Solanum pennellii**. Nature Genetics 2014, 46, 1034-1038.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F: **Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature.** PLoS Genet 2010, 6(5):e1000940.

Bucheli P, Voirol E, de la Torre R, López J, Rytz A, Tanksley SD, Pétiard V: **Definition of nonvolatile markers for flavor of tomato (*Lycopersicon esculentum* Mill.) as Tools in Selection and Breeding.** Journal of Agricultural and Food Chemistry 1999, 47, 659–664. Goff y Klee 2006

Busch BL, Schmitz G, Rossmann S, Piron F, Ding J, Bendahmane A, Theres K: **Shoot branching and leaf dissection in tomato are regulated by homologous gene modules.** The Plant Cell 2011, 23, 3595–3609.

Carrari F & Fernie AR: **Metabolic regulation underlying tomato fruit development.** Journal of Experimental Botany 2006, 57(9), 1883–1897. doi:10.1093/jxb/erj020

Cassera MB, Gozzo FC y col: **The methylerythritol phosphate pathway is functionally active in all intraerythrocytic stages of *Plasmodium falciparum*.** J. Biol. Chem. 2004, 279, 51749–51759.

Causse M, Saliba-Colombani V, Lesschaeve I, Buret M: **Genetic analysis of organoleptic quality in fresh market tomato.** Theoretical and Applied Genetics 2001, 102, 273–283.

Causse M, Saliba-Colombani V, Lecomte L, Duffé P, Rousselle P, Buret M: **QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits.** J Exp Bot 2002, 53:2089–2098

Causse M: **Whole genome resequencing in tomato reveals variation associated with introgression and breeding events.** BMC Genomics 2013, 14:14:791.

Chander S y col: **Genetic dissection of tocopherol content and composition in maize grain using quantitative trait loci analysis and the candidate gene approach.** Mol. Breed. 2008, 22, 353–365.

Changtian Pan, Lei Ye, Li Qin, Xue Liu, Yanjun He, Jie Wang, Lifei Chen & Gang Lu: **CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening.** Scientific Reports 2016.doi:10.1038/srep24765

Chen Z: **Genomic and epigenetic insights into the molecular bases of heterosis.** Nat Rev Genet 2013, 14: 471-482.

Diepenbrock C, Kandianis C y col: **Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain** 2017. DOI: <https://doi.org/10.1105/tpc.17.00475>

Chung S, & Felton G: **Specificity of Induced Resistance in Tomato Against Specialist Lepidopteran and Coleopteran Species.** Journal of Chemical Ecology 2011, 37(4), 378–386. doi:10.1007/s10886-011-9937-0 Gebhardt y Valkonen 2001

Coley P, Bryant J, Chapin Fr: **Resource availability and plant antiherbivore defense.** Science 1985, 230:895.12.

Conklin PL, Saracco SA, Norris R & Last RL: **Identification of Ascorbic Acid Deficient Arabidopsis thaliana Mutants**. Genetics 2000, 154 (2), 847–856.

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, Caillieux E, Hospital F, Aury JM, Wincker P, Roudier F, Jansen RC, Colot V, Johannes F: **Mapping the epigenetic basis of complex traits**. Science 2014. 7;343(6175):11458.

Cubas P, Vincent C. y Coen E: **An epigenetic mutation responsible for natural variation in floral symmetry** Nature 1999, 157–161 doi:10.1038/43657

Daniel A. Griffith P y col: **The Box-Cox Transformation: Computational and Interpretation Features of the Parameters**.

Davenport CB: **Recessive characters**. Science 1908,20; 28(725):729.

Davuluri GR, Tuinen A, Van Fraser PD, Manfredonia A, Newman R, Burgess D, Brummell DA y col: **Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes**. Nature Biotechnology 2005,23(7).doi:10.1038/nbt1108

De Jong M, Mariani C & Vriezen WH: **The role of auxin and gibberellin in tomato fruit set**. Journal of Experimental Botany 2009, 60(5), 1523–1532. doi:10.1093/jxb/erp094

DellaPenna D y Pogson B: **Vitamin synthesis in plants: tocopherols and carotenoids**. Annual review of plant biology 2006 ,57,711–38. doi:10.1146/annurev.arplant.56.032604.144301

Dixon RA and Strack D: **Phytochemistry meets genome analysis, and beyond**. Phytochemistry 2003, 62:815-16

Do PT, Prudent M, Sulpice R, Causse M, Fernie AR: **The influence of fruit load on the tomato pericarp metabolome in a Solanum chmielewskii introgression line population**. Plant Physiol 2010, 154:1128-42

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker: **Fitting Linear Mixed-Effects Models Using lme4**. Journal of Statistical Software 2015, 67(1), 1-48. doi:10.18637/jss.v067.i01

Downen RH, Pelizzola M, Schmitz RJ , Lister R, Downen JM, Nery JR, Dixon JE y col: **Widespread dynamic DNA methylation in response to biotic stress**. Proc. Natl. Acad. Sci. 2012, USA 109, E2183–E2191.

Dubin MJ, Zhang P, Meng D, Remigereau MS, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G y col: **DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation**. eLife 2015, e05255.

Enfissi E, Fraser D, Lois LM, Boronat A, Schuch W., & Bramley P. M. **Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate forming pathways for the production of health-promoting isoprenoids in tomato**. Plant Biotechnology Journal 2005, 3(1), 17–27. doi:10.1111/j.1467-7652.2004.0009

Eshed Y, Zamir D: **An Introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL.** Genetics 1995, 141, 1147–1162.

Evans HM, Bishop KS: **"On the existence of a hitherto unrecognized dietary factor essential for reproduction".** Science 1922, 56 (1458): 650–651.

Falk J & Munné-Bosch S: **Tocochromanol functions in plants: antioxidation and beyond.** Journal of Experimental Botany 2010, 61(6), 1549–1566. doi:10.1093/jxb/erq030

FAO Water: **Crop Water Information Soybean.** 2015, http://www.fao.org/nr/water/cropinfo_tomato.html.

Feng F, Feng D y col: **QTL mapping for the tocopherols at milk stage of kernel development in sweet corn.** Euphytica 2013, Vol. 193 Issue 3, p409-417

Fernie AR y Tohge T: **The Genetics of Plant Metabolism.** Annu.Genet. 2017, 51:287–310, <https://doi.org/10.1146/annurev-genet-120116-024640>

Fernie AR: **The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding.** Phytochemistry 2007, 68:2861-80

Fernie AR & Willmitzer L: **Molecular and Biochemical Triggers of Potato Tuber Development.** Plant Physiolog 2001, 127(4), 1459–1465. doi:10.1104/pp.010764

Fitzpatrick T y col: **Vitamin deficiencies in humans: Can Plant Science help?** Plant Cell 2012, 24 (2):395-414.

Flint J, Mackay TF: **Genetic architecture of quantitative traits in mice, flies, and humans.** Genome Res 2009, 19(5):723-733

Foolad MR: **Genome mapping and molecular breeding of tomato.** Int J Plant Genomics. 2007, 64358. doi: 10.1155/2007/64358

Fraser PD, Enfissi EM, Halket JM, Truesdale MR, Yu D, Gerrish C, Bramley PM: **Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism.** The Plant cell 2007, 19(10), 3194–211. doi:10.1105/tpc.106.049817

Fraser P, Pint M, Holloway D, Bramley P: **Application of high-performance liquid chromatography with photodiode array detection to the metabolite profiling of plant isoprenoids.** The Plant Journal 2000, (24), 551–558.

Fray RG y Grierson D: **Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression.** Plant Molecular Biology 1993, 22: 589–602

Fritsche S, Wang X, Li J, Stich B, Kopisch obuch FJ, Leckband G, Dreyer F, y col: **A candidate gene-based association study of tocopherol content and composition in rapeseed (*Brassica napus*).** Frontiers in Plant Science 2012, 1–24. doi:10.3389/fpls.2012.00129

Fulton TM, Nelson JC, Tanksley SD: **Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, into *Lycopersicon esculentum*, followed through three successive backcross generations.** Theoretical and Applied Genetics October 1997, 895–902

Fulton TM, Grandillo S, Beck Bunn T, Fridman E, Frampton A, Lopez J, y col: **Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross.** Theoretical and Applied Genetics , 2000, 100:1025-1042

Furuya T, Yoshikawa T, Kimura T, Kaneko H. **Production of tocopherols by cell culture of safflower.** Phytochemistry 1987, 26, 2741–2747.

Futuyma DJ y Agrawal A. **Macroevolution and the biological diversity of plants and herbivores.** PNAS 2009, 106:18054-61

Georgiadou C, Goulas V y col: **Regulation of On-Tree Vitamin E Biosynthesis in Olive Fruit during Successive Growing Years.** Plant Sci., 16 November 2016. doi.org/10.3389/fpls.2016.01656

Georgiadou C, Thessaloniki Ntourou², Vlasios Goulas y col: **Temporal analysis reveals a key role for VTE5 in vitamin E biosynthesis in olive fruit during on-tree development.** Plant Sci., 21 Front. 2015 doi.org/10.3389/fpls.2015.00871

Farré G, Sudhakar D, Naqvi S, Sandmann G, Christou P y col: **Transgenic rice grains expressing a heterologous p-hydroxyphenylpyruvate dioxygenase shift tocopherol synthesis from the γ to the α isoform without increasing absolute tocopherol levels.** Transgenic Research October 2012, Volume 21, Issue 5, pp 1093–1097

Gibson G: **Rare and common variants: twenty arguments.** Nat Rev Genet 2011, 13(2):135-145.

Gilliland LU, Magallanes-Lundback M, Hemming C, Supplee A, Koornneef M, Bentsink L, DellaPenna D: **Genetic basis for natural variation in seed vitamin E levels in *Arabidopsis thaliana*.** Proc. Natl. Acad. Sci. USA 2006, 103, 18834–18841

Giovannoni J: **Molecular biology of fruit maturation and ripening.** Annu Rev Plant Physiol Plant Mol Biol 2001, 52: 725–749

Giovannoni J: **Genetic Regulation of Fruit Development and Ripening.** The Plant Cell. 2004, 16:1, 170–180. doi:10.1105/tpc.019158

Giovinazzo G, D’Amico L, Paradiso A, Bollini R, Sparvoli F, De Gara L: **Antioxidant metabolite profiles in tomato fruit constitutively expressing the grapevine stilbene synthase gene.** Plant Biotechnology Journal 2005, 3(1),57–69. doi:10.1111/j.1467-7652.2004.00099.x

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK: **ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase.** Science. 2002, 12;296(5566):343-6

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J: **qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data**. *Genome Biol* 8 *Genome biology*, 2007, 8, 19

Hong W, Shutu X, Yaming F, Nannan L, Wei Zhany col: **Beyond pathways: genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses**. 2018 DOI: 10.1111/pbi.12889

Huang B y George A: **R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines**. *Bioinformatics* 2011, Mar 1;27(5):727-9. doi:10.1093/bioinformatics/btq719.

Xu J, Ranc N, Muños S y col: **Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species**, *Theor. Appl. Genet.* 126 2013, 567–581.

Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC and Gao G. **PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants**. *Nucleic Acids Research* 2017, 45(D1):D1040-D1045

Kamenetzky L y col: **Genomic analysis of wild tomato (*Solanum pennellii*) introgressions determining metabolic- and yield-associated traits**. *Plant Physiol* 2010, 152(4):1772-86.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E: **Efficient control of population structure in model organism association mapping**. *Genetics* 2008, 178:1709–1723 (Chen y Chen 2008).

Kawashima T and Berger F: **Epigenetic reprogramming in plant sexual reproduction**. *Nat. Rev. Genet* 2014; 15, 613–624

Keene JD: **RNA regulons: coordination of post-transcriptional events**. *Nat. Rev. Genet* 2007, 8: 533-543.

Keurentjes J and Fernie AR: **Genetics, genomics and metabolomics**. In *Annual Plant Reviews* 2011, pp 239-52.

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D y col: **Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation**. *Plant Physiol* 2001. 126:811-25

Kobayashi N and DellaPenna D: **Tocopherol metabolism, oxidation and recycling under high light stress in *Arabidopsis***. *Plant J.* 2008, 55(4):607-18. doi: 10.1111/j.1365-313X.2008.03539.x.

Kolomiets MV, Hannapel DJ, Chen H, Tymeson M & Gladon RJ: **Lipoxygenase Is Involved in the Control of Potato Tuber Development**. *The Komili S, Silver PA* (2008). 124:322-43

Korte A y Farlow A: **The advantages and limitations of trait analysis with GWAS**. *Plant Methods* 2013, 9:29 <http://www.plantmethods.com/content/9/1/29>

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R: **A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana***. PLoS Genet 2009, 5(7):e1000551

Krieger-Liszkay, A., & Trebst: **A Tocopherol is the scavenger of singlet oxygen produced by the triplet states of chlorophyll in the PSII reaction centre**. Journal of experimental botany 2006, 57(8), 1677–84. doi:10.1093/jxb/erl002

Kriventseva EV, Tegenfeldt F y col: **"OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software"**. Nucleic Acids 2015, D250–6. doi:10.1093/nar/gku1220. PMC 4383991

Kruk J & Trebst A: **Plastoquinol as a singlet oxygen scavenger in photosystem II**. Bioenergetics 2008: 1777(2), 154–162. doi:http://dx.doi.org/10.1016/j.bbabi.2007.10.008

Kruk J, Holländer-Czytko H, Oettmeier W, & Trebst A: **Tocopherol as singlet oxygen scavenger in photosystem II**. Journal of Plant Physiology 2005, 162(7), 749–757. doi:http://dx.doi.org/10.1016/j.jplph.2005.04.020

Lecomte L, Saliba-Colombani V y col: **Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato**, Mol. Breed 2004, 1–14.

Li DP, Baldwin IT, Gaquerel E: **Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis**. PNAS 2015, 112:E4147-55

Li H, Wang Y, Han Y, Teng W y col: **Mapping quantitative trait loci (QTLs) underlying seed vitamin E content in soybean with main, epistatic and QTL environment effects**. Plant Breed. 2016, 135, 208–214.

Lipka, A.E.; Gore, M.A.; Magallanes-Lundback, M.; Mesberg, A.; Lin, H.; Tiede, T.; Chen, C.; Buell, C.R.; Buckler, E.S.; Rocheford, T.; et al. **Genome-wide association study and pathway-level analysis of tocopherol levels in maize grain**. G3 (Bethesda) 2013, 3, 1287–1299.

Lisch D: **Epigenetic regulation of transposable elements in plants**. Annu Rev Plant Biol. 2009, 60:43-66. doi: 10.1146/annurev.arplant.59.032607.092744.

Loudet O, Chaillou S, Merigout P, Talbotec J, Daniel-Vedele F: **Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis***. Plant Physiol 2003, 131:345-58

Guyer L y col: **A cytochrome P450 regulates a domestication trait in cultivated tomato**, Proc. Natl. Acad. Sci. U.S.A. 110 2013, 17125–17130.

Mach J: **Phytol from degradation of Chlorophyll feeds biosynthesis of Tocopherols**. Plant Cell 2015, 27(10):2676.

Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A: **The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators**. Plant Physiol 2008, 148:2021-2049.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB: **A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening.** Nat Genet 2006, 38, 948–952.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A y col: **Finding the missing heritability of complex diseases.** Nature 2009, 461(7265):747-753.

Maruyama-Nakashita A, Nakamura Y, Tohge T, Saito K, Takahashi H: **Arabidopsis SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism.** Plant Cell 2006, 18:3235-3251.

Matt P, Krapp A, Haake V, Mock H, Stitt M: **Decreased Rubiscoactivity leads to dramatic changes of nitrate metabolism, amino acid metabolism and the levels of phenylpropanoids and nicotine in tobacco antisense RBCS transformants.** Plant J 2002, 30:663-677.

Matzke and Matzke: **Planting the Seeds of a New Paradigm Marjori.** PLoS Biol. 2004, 2(5): e133. Published online 2004 May 11. doi: 10.1371/journal.pbio.0020133 PMID: PMC406394

Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ: **RNA-directed DNA methylation and Pol IVb in Arabidopsis.** Cold Spring Harb Symp Quant Biol 2006, 71:449-59.

Miller JC, Tanksley SD: **RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon.** Theor Appl Genet. 1990;80:437–48

Miret, J. and Munné-Bosch: **Redox signaling and stress tolerance in plants: a focus on vitamin E.** Ann. N. Y. Acad. Sci. 2015, 1–10.

Mitchell-Olds T, Pedersen D: **The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis.** Genetics 2003, 149:739-47

Moco, S., Capanoglu, E., Tikunov, Y., Bino, R. J., Boyacioglu, D., Hall, R. D., Vervoort y col: **Tissue specialization at the metabolite level is perceived during the development of tomato fruit.** Journal of experimental botany 2007, 58(15-16), 4131–46. doi:10.1093/jxb/erm271

Moore y col: **A linear mixed model approach to study multivariate geneenvironment interactions.** bioRxiv 2018, 10.1101/270611

Muir SR, Collins GJ y col: **Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols.** Nature biotechnology 2001, 19(5), 470–4. doi:10.1038/88150

Munné-Bosch S & Alegre L: **Interplay between ascorbic acid and lipophilic antioxidant defences in chloroplasts of water-stressed Arabidopsis plants.** FEBS letters 2002, 524(1-3), 145–8.

Napoli C, Lemieux C, Jorgensen R: **Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans.** Plant Cell. 1990, Apr;2(4):279-289.

Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ **Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids.** *Nature* 2009, 457, 327-331.

Oliver RE, Islamovic E y col.: **Comparative systems biology reveals allelic variation modulating tocochromanol profiles in barley (*Hordeum vulgare* L.).** *PLoS ONE* 2014, 9, e96276.

Haddadi P y col: **Genetic dissection of tocopherol and phytosterol in recombinant inbred lines of sunflower through quantitative trait locus analysis and the candidate gene approach.** *Molecular Breeding* 2012, 29: 3, pp 717–729

Park YH, West MAL, St Clair DA: **Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L).** *Genome*. 2004; 47:510–8

Pascual L y col: **Deciphering genetic diversity and inheritance of tomato fruit weight and composition through a systems biology approach.** *J. Exp Bot* 2013, 64(18):5737-52.

Pascual, L. y col: **Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era.** *Plant Biotechnol J* 2015, 13: 565-577. doi:10.1111/pbi.12282

Pellaud S. and Mène Saffrané L: **Metabolic Origins and Transport of Vitamin E Biosynthetic Precursors.** *Front. Plant Sci.* 2017: 8, 1959.

Peralta I, Spooner D, & Knapp S: **Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *juglandifolia*, sect. *Lycopersicon*; *Solanaceae*).** *Systematic Botany Monographs* 2008, 84, 1–186.

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez JP, Ganai M, Zamir D, Lifschitz E: **The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1.** *Development* Cambridge, England 1998, 125, 1979–1989.

Pouvreau B y col: **Duplicate maize Wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis.** *Plant Physiol.* 2011, 156, 674–686.

Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, Conti G, Corrêa da Silva JV, Peralta IE, Colot V, Asurmendi S, Fernie AR, Rossi M, Carrari F: **Natural occurring epialleles determine vitamin E accumulation in tomato fruits.** *Nature Commun.* 2014, 26; 5:3027. doi: 10.1038/ncomms5027.

Quadrana L, Almeida J, Otaiza S, Duffy T, Corrêa da Silva JV, de Godoy F, Asís R, Bermúdez L, Fernie AR, Carrari F, Rossi M: **Transcriptional regulation of tocopherol biosynthesis in tomato.** *Plant Mol. Biol* 2012.

Ranc N, Muños S, Santoni S, Causse M: **A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (*Solanaceae*).** *BMC Plant Biol* 2008, 8:130

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J. 2000. **An alternative pathway to β -carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato.** Pnas 97, 11102–11107

Rossi M, Bermúdez L, Carrari F: **Crop yield: challenges from a metabolic perspective.** Curr Opin Plant Biol 2015, 20;25:79-89.

Rossi M, Goggin FL, Milligan SB y col: **The nematode resistance gene Mi of tomato confers resistance against the potato aphid.** Proceedings of the National Academy of Sciences of the United States of America 1998, 95(17), 9750–9754. doi:10.1073/pnas.95.17.9750

Ruijter JM, Ramakers C, Hoogaars WMH y col: **Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data.** Nucleic acids research 2009, 37(6), e45. doi:10.1093/nar/gkp045

Sim s, Van Deynze A y col: **High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding,** PLoS ONE 7 (2012) e45520.

Sacco A, Ruggieri V, Parisi M, Festa G, Rigano MM, Picarella ME, Mazzucato A, Barone A: **Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform.** PLoS ONE 2015, 10, 1–20

Saito K, Tautz L. and Mustelin T: **The lipid-binding SEC14 domain.** Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 2007, 1771, 719–726

Saldeen P & Saldeen T: **Women and omega-3 Fatty acids.** Obstet Gynecol Surv 2004, 59, 722–730. Terasawa y col. 2000

Sato T, Iwatsubo T, Takahashi M, Nakagawa H, Ogura N, Mori H. 1993. **Intercellular localization of acid invertase in tomato fruit and molecular cloning of a cDNA for the enzyme.** Plant & cell physiology 34, 263–9. (Schumacher y col. 1999)

Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M. **GenomeWide Association in tomato reveals 44 candidate loci for fruit metabolic traits.** Plant physiology 2014, 165, 1120–1132

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D y col: **Mode of inheritance of primary metabolic traits in tomato.** Plant Cell 2008, 20:509-23

Schauer, N., Zamir, D., & Fernie, A. R. **Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex.** Journal of Experimental Botany 2005 56(410), 297–307. doi:10.1093/jxb/eri057 Blanca y col. 2015

Schauer N, Semel Y, Roessner U y col: **Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement.** Nature biotechnology 2006, 24(4), 447–54. doi:10.1038/nbt1192

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K y col: **GC-MS libraries for the rapid identification of metabolites in complex biological samples.** FEBS letters 2005, 579, 1332–1337.

Schillmiller AL, Shi F, Kim J, Charbonneau AL, Holmes D y col: **Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines.** Plant J 2010, 62:391-403

Schwarz, Gideon E: **Estimating the dimension of a model**». Annals of Statistics 1978, (2): 461–464. MR 468014. doi:10.1214/aos/1176344136.

Secco D, Wang C, Shou H, Schultz, MD, Chiarenza S, Nussaume y col: **induced gene expression drives transient DNA methylation changes at adjacent repetitive elements.** eLife 2015

Segura V, Vilhjalmsón BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M: **An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations.** Nat Genet 2012, 44 (7):825-830. 40.

Seo YS, Kim SK, Harn CH, Kim WT: **Ectopic expression of apple fruit homogenisate phytyltransferase gene (MdHPT1) increases tocopherol in transgenic tomato (Solanum lycopersicum cv. Micro-Tom) leaves and fruits.** Phytochemistry 2011, 72: 321–329

Cokus S y col: **Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.** Nature 2008, doi:10.1038/nature06745

Sim S-C, Durstewitz G, Plieske J, et al.. **Development of a large SNP genotyping array and generation of high-density genetic maps in tomato.** PloS one 2012, 7, e40563.

Soll J y Schultz G: **Comparison of geranylgeranyl and phytyl substituted methylquinols in the tocopherol synthesis of spinach chloroplasts.** Biochemical and Biophysical Research Communications 1979, 91(3), 715–720. doi:http://dx.doi.org/10.1016/0006-291X(79)91939-9

Sonderby I, Hansen B, Bjarnholt N, Ticconi C, Halkier B, Kliebenstein D: **A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates.** PLoS ONE 2007, 2:e1322.

Spicher L and Kessler F: **Unexpected roles of plastoglobules (plastid lipid droplets) in vitamin and E metabolism.** Curr. Opin. Plant Biol 2015, 25, 123–129.

Stegmayer G: **A Biologically-inspired Validity Measure for Comparison of Clustering Methods over Metabolic Datasets.** IEEE/ACM Trans. Comput. Biol. Bioinf 2012, 9 (3): 706-716.

Steinhauser MC: **Enzyme Activity Profiles during Fruit Development in Tomato Cultivars and Solanum pennellii.** Plant Physiol 2010, 153(1):80-98.

Stevens R, Buret M, Garchery C, Carretero Y, Causse M: **Technique for Rapid, Small-Scale Analysis of Vitamin C Levels in Fruit and Application to a Tomato Mutant collection.** Journal of Agricultural and Food Chemistry 2006, 54(17),6159–6165. doi:10.1021/jf061241e

Strassman M, Ceci LN: **Enzymatic formation of alphaisopropylmalic acid, an intermediate in leucine biosynthesis.** J Biol Chem 1963, 238:2445-2452.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G y col: **Genetic determinants of hair, eye and skin pigmentation in Europeans**. Nat Genet 2007, 39(12):1443-1452.

Sun Q, Zybaylov B, Majeran W, Olinares P, Van Wijk: **KPPDB, the Plant Proteomics Database at Cornell**. Nucleic acids research 2009, 37, 969–974.

Sussmann R, Angeli C, Peres V y col: **Intraerythrocytic stages of Plasmodium falciparum biosynthesize vitamin E**. FEBS Lett. 2011, 585, 3985–3991.

Swamy B and Sarla N: **Yield enhancing quantitative trait loci (QTLs) from wild species**. Biotechnology Advances. 2008;26(1):106–120.doi: 10.1016/j.biotechadv.2007.09.005

Kawakatsu T y col: **Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions**, 2010 <http://dx.doi.org/10.1016/j.cell.2016.06.044>

Tang Y, Fu X., Shen Q., Tang K: **Roles of MPBQ-MT in promoting alpha/gamma-tocopherol production and photosynthesis under high light in lettuce**. PLoS ONE 2016, 11:e0148490. [10.1371/journal.pone.0148490](https://doi.org/10.1371/journal.pone.0148490)

The Tomato Genome Consortium: 312 authors. **Tomato genome sequencing and comparative analysis reveal two consecutive triplications that spawned genes influencing fruit characteristics**. Nature 2012, 485:635-641.

Thompson AJ, Jackson AC, Parker RA, Morpeth DR, Burbidge A, Taylor IB: **Absciscic acid biosynthesis in tomato: Regulation of zeaxanthin epoxidase and 9-cis-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid**. Plant Molecular Biology 2000, 42, 833–845

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, Mc Mullen MD, Holland JB, Buckler ES: **Genome-wide association study of leaf architecture in the maize nested association mapping population**. Nat Genet 2011, 43(2):159–162

Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P y col: **Identification of loci affecting flavour volatile emissions in tomato fruits**. J. Exp. Bot 2006, 57:887-96

Tieman DM, Zhu G, Resende MF, Lin T, Nguyen C: **A chemical genetic roadmap to improved tomato flavor**. Science 2017, 355:391–94

Traber MG y Atkinson J: **Vitamin E, antioxidant and nothing more**. Radic Biol Med. 2007, 1;43(1):4-15

Traber, M. G., & Sies, H. (1996). **Vitamin E in Humans: Demand and Delivery**. Annual Review of Nutrition 1996, 16(1), 317–321. doi:10.1146/annurev.nu.16.070196.001541

Tzin V, Malitsky S, Aharoni A, Galili G: **Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in Arabidopsis**. Plant J. 2009, 60, 156–167.

Valentin H, Lincoln K, Moshiri F y col: **The Arabidopsis vitamin E pathway gene5-1 mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis.** Plant Cell 2006, 18, 212–224.

Van der Rest B, Danoun S, Boudet AM, & Rochange, SF: **Down-regulation of cinnamoyl-CoA reductase in tomato (*Solanum lycopersicum* L.) induces dramatic changes in soluble phenolic pools.** Journal of Experimental Botany 2006: 57(6), 1399–1411. doi:10.1093/jxb/erj120

Vera-Ruiz E, Velasco A y col: **Genetic Mapping of the Tph1 Gene Controlling Beta-tocopherol Accumulation in Sunflower Seeds QTL Associated with Accumulation of Tocopherols in Maize.** Crop Science 2003, DOI: 10.2135/cropsci2003.2257

Vidi PA, Kessler F y Bréhélin C: **Plastoglobules: a new address for targeting recombinant proteins in the chloroplast.** BioMed Central 2007, <https://doi.org/10.1186/1472-6750-7-4>

Vom Dorp, K.; Hölzl, G.; Plohm, C.; Eisenhut, M.; Abraham, M.; Weber, A.P.M.; Hanson, A.D.; Dörmann, P. **Remobilization of phytol from chlorophyll degradation is essential for tocopherol synthesis and growth of Arabidopsis.** Plant Cell 2015, 27, 2846–2859.

Soppe y col: **The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene.** Molecular Cell 2000, Pages 791–802 [https://doi.org/10.1016/S1097-2765\(05\)00090-0](https://doi.org/10.1016/S1097-2765(05)00090-0)

Wang X, Zhang C, Li L, Steffi Fritsche y col: **Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (*Brassica napus* L.)** November 20, 2012 <https://doi.org/10.1371/journal.pone.0050038>

Wang Y, van der Hoeven RS, Nielsen R, Mueller LA, Tanksley SD: **Characteristics of the tomato nuclear genome as determined by sequencing undermethylated EcoRI digested fragments.** Theor. Appl. Genet., 2005. 112(1): p. 72–84

Wang Y, van der Hoeven RS, Nielsen R, Mueller LA, Tanksley SD: **Characteristics of the tomato nuclear genome as determined by sequencing undermethylated EcoRI digested fragments.** Theor. Appl. Genet., 2005. 112(1): p. 72–84

Wang M, Toda K, Maeda HA: **Biochemical properties and subcellular localization of tyrosineaminotransferases in Arabidopsis thaliana.** Phytochemistry 2016, 132, 16–25.

Wang X, Zhang C, Li L y col: **Unraveling the genetic basis of seed tocopherol content and composition in rapeseed (*Brassica napus* L.).** PLoS ONE 2012, 7, e50038.

Warnes G, Leisch F: **Genetics: Population Genetics: 2005. R package version 1.2.0.**

Wassenegger M. **RNA-directed DNA methylation.** Plant Mol. Biol. 2000, 43:203–20

Wassenegger M, Heimes S, Riedel L, Sängler HL: **RNA-directed de novo methylation of genomic sequences in plants** Cell 1994, 76(3):567–76.

Weller JL, Perrotta G, Schreuder MEL, Van Tuinen A, Koornneef M, Giuliano G, Kendrick RE: **Genetic dissection of blue-light sensing in tomato using mutants deficient in cryptochrome 1 and phytochromes A, B1 and B2.** The Plant Journal 2001, 25, 427–440.

Wilkinson JQ, Lanahan MB, Yen H-C, Giovannoni JJ, Klee HJ: **An ethylene-inducible component of signal transduction encoded by never-ripe.** Science 1995, 270, 1807–1809.

Wolf G: **History of Nutrition The Discovery of the Antioxidant Function of Vitamin E: the Contribution of Henry A. Mattill.** The Journal of Nutrition 2005, 135, 363–366.

Xu J y col: **An extensive proteome map of tomato (Solanum lycopersicum) fruit pericarp.** Proteomics 2013, 13(20):3059-63.

Yan Y, Yuan Z, Yingjun C, Baofang F & Zhixiang C: **Characterization of Soybean WRKY Gene Family and Identification of Soybean WRKY Genes that Promote Resistance to Soybean Cyst Nematode.** Scientific Reports 2017, 17804 doi:10.1038/s41598-017-18235-8

Yokota T, Igarashi K, Uchihara T, Jishage K, Tomita H, Inaba A, Li Y: **Delayed-onset ataxia in mice lacking β -tocopherol transfer protein: Model for neuronal degeneration caused by chronic oxidative stress.** Proceedings of the National Academy of Sciences 2001, 98 (26): 15185–15190. doi:10.1073/pnas.261456098

Ytterberg A , Peltier J, Wijk K Van: **Protein Profiling of Plastoglobules in Chloroplasts and Chromoplasts . A Surprising Site for Differential Accumulation of Metabolic Enzymes.** Plant physiology 2006, 140, 984–997.

Zanor MI, Rambla JL, Chaib J, Steppa A, Medina A, y col: **Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents.** J. Exp Bot. 2009, 60:2139-54

Zeng P, Zhao Y, Qian C, Zhang L, Zhang R y col: **Statistical analysis for genome-wide association study.** J. Biomed. Res 2015, 29:285-97

Zhang J, Chen R, Xiao J, Qian C, Wang T, Li H, Ouyang B, Ye Z. **A single-base deletion mutation in SlIAA9 gene causes tomato (Solanum lycopersicum) entire mutant.** Journal of Plant Research 2007, 120, 671–678.

Zhang C, Cahoon RE, Hunter SC, Chen M, Han J, Cahoon EB: **Genetic and biochemical basis for alternative routes of tocotrienol biosynthesis for enhanced vitamin E antioxidant production.** Plant J. 2013, 73, 628–639.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P: **An Arabidopsis example of association mapping in structured samples.** PLoS Genet 2007, 3(1):e4.

Zhaobo Lang, Yihai Wang, Kai Tang, Dengguo Tang, Tatsiana Datsenka, Jingfei Cheng, Yijing Zhang, Avtar K. Handa and Jian-Kang Zhu: **Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit.** PNAS 2017, 114 (22) E4511-E4519. <https://doi.org/10.1073/pnas.1705233114>

Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, Shao Y, Giovannoni JJ: **Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening.** Nat Biotechnology 2013, 31(2):154.

Zou L, Li H, Ouyang B, Zhang J, & Ye Z: **Molecular cloning, expression and mapping analysis of a novel cytosolic ascorbate peroxidase gene from tomato.** DNA Seq. 2005, 16, 456–61.

Anexo

Tabla S1. Visión global del conjunto de materiales utilizados en esta tesis a través de los distintos capítulos



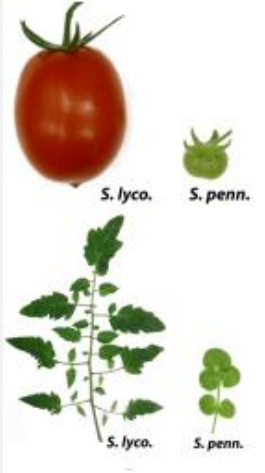
Capítulo	CAPITULO II	CAPITULO II	CAPITULO III
Materiales	<p>Población de GWAS (146 accesiones)</p> 	<p>F7 a partir de 8 líneas parentales MAGICPOP (124 líneas)</p> 	<p>F2 cruza (IL9-2-6-1 X IL3-1) (80 líneas recombinantes)</p> 
Datos genotípicos	6100 SNPs (Sim y col. 2012)	1345 Marcadores (Pascual y col. 2015)	Marcadores alelo específicos asociados a los genes parálogos VTE3(1 y 2)
Ensayos	146 accesiones en dos condiciones de estrés hídrico y control	124 líneas y sus parentales en condiciones de invernáculo	80 líneas recombinantes F2 y sus parentales en condiciones de campo e invernáculo
Objetivo científico	Búsqueda de nuevos loci asociados VTE	Búsqueda de nuevos loci asociados a VTE	Estudio de epistasis entre los parálogos VTE3(1y2)

Tabla S2. Materiales GWAS: banco de germoplasma; origen geográfico; grupo genético; sub-grupo genético; referencias; color de fruto; contenidos de tococromanoles en frutos

Genotipo ⁽¹⁾	Banco de germoplasma ⁽²⁾	Origen geográfico ⁽³⁾	Grupo genético ⁽⁴⁾	Sub-grupo genético ⁽⁵⁾	(Albert y col. 2017) ⁽⁶⁾	(Blanca y col. 2015) ⁽⁷⁾	(Sauvage y col. 2015, Pascual y col. 2015) ⁽⁸⁾	Color de fruto	Contenidos VTE frutos ⁽⁹⁾											
									α-Tocoferol [μg/gPS]		α-Tocotrienol [μg/gPS]		β-Tocoferol [μg/gPS]		γ-tocoferol [μg/gPS]		δ-tocoferol [μg/gPS]		Total Tocoferol [μg/gPS]	
									C	E	C	E	C	E	C	E	C	E	C	E
BGV12627	COMAV	Colombia	SLC	SLC_no_Andino	SI	SI	NO	rojo	66.1	53.4	12.9	10.2	5.1	5.4	13.0	7.5	3.7	1.6	88.0	67.9
BGV13157	COMAV	Perú	SLC	SLC_Perú	SI	SI	NO	rojo	68.4	97.4	16.2	21.5	9.5	8.8	14.7	27.4	0.7	4.8	93.4	138.3
BGV13161	COMAV	Perú	SLC	SLC_Perú	SI	SI	NO	NA	81.3	64.6	18.3	13.2	8.5	5.9	17.6	12.4	3.1	1.9	110.5	84.9
BGV13945	COMAV	Perú	SLC	SLC_Perú	SI	SI	NO	rojo	42.1	29.0	19.5	14.1	3.2	1.9	38.0	35.7	4.2	4.1	87.5	70.7
BGV5895	COMAV	Bolivia	SLC	SLC_no_Andino	SI	SI	NO	rojo	50.1	75.4	11.2	13.5	6.5	4.2	12.6	11.5	0.5	3.0	69.7	94.1
BGV6071	COMAV	Ecuador	SLC	SLC_no_Andino	SI	SI	NO	rojo	30.4	95.9	15.5	7.0	3.7	1.2	13.7	4.7	4.1	4.4	51.9	106.3
BGV6767	COMAV	Ecuador	SLC	SLC_Ecuador	SI	SI	NO	rojo	56.2	77.6	0.8	1.1	1.5	1.6	15.3	23.4	0.7	1.4	73.7	104.0
BGV6769	COMAV	Ecuador	SLC	SLC_Ecuador	SI	SI	NO	rojo	73.3	63.1	18.1	14.3	6.1	7.2	11.6	13.4	3.5	0.6	94.4	84.4
BGV6779	COMAV	Ecuador	SLC	SLC_Ecuador	SI	SI	NO	rojo	56.1	56.0	16.3	14.8	1.9	3.9	9.6	10.4	5.8	4.2	73.5	74.4
BGV6827	COMAV	Ecuador	SLC	SLC_Ecuador	SI	SI	NO	rojo	53.5	50.8	16.1	12.7	4.5	3.1	20.1	18.9	4.1	3.8	82.2	76.7
BGV6931	COMAV	Ecuador	SLC	SLC_Ecuador	SI	SI	NO	rojo	73.1	70.7	18.0	13.1	1.8	5.1	14.9	9.1	6.3	2.2	96.1	87.1
BGV7888	NA	NA	NA	NA	NO	NO	NO	NA	101.4	88.8	2.1	4.7	2.6	2.6	22.1	18.0	0.5	1.6	126.6	111.1
BGV7942	COMAV	Cuba	SLC	SLC_no_Andino	SI	SI	NO	rojo	83.2	89.9	0.3	2.5	2.7	2.6	27.1	17.8	1.6	2.7	114.7	113.0
BGV7979	COMAV	Perú	SLC	SLC_Perú	SI	SI	NO	rojo	55.4	46.8	2.1	6.8	1.5	1.2	48.0	38.8	2.6	2.4	107.5	89.2
BGV8036	COMAV	Bolivia	SLC	SLC_Perú	SI	SI	NO	rojo	76.3	88.6	1.6	2.6	1.5	1.8	14.8	15.3	0.4	1.0	93.0	106.8
BGV8096	COMAV	Perú	SLC	SLC_Perú	SI	SI	NO	rojo	105.7	67.6	4.0	6.1	4.8	2.0	31.1	16.2	1.2	2.1	142.8	87.9
BGV8109	COMAV	Perú	SP	SLC_SP_Perú	SI	SI	NO	rojo	28.3	40.2	14.7	8.6	3.6	2.4	28.5	30.3	3.2	2.3	63.6	75.1
BGV8219	COMAV	Costa Rica	SLC	SLC_no_Andino	SI	SI	NO	rojo	70.3	79.1	1.3	3.5	1.4	1.0	9.3	4.1	0.4	1.2	81.5	85.4
BGV8221	COMAV	Salvador	SLC	SLC_no_Andino	SI	SI	NO	rojo	67.7	111.4	3.5	1.5	1.8	2.8	17.6	9.3	1.0	1.4	88.0	124.9
BGV8228	COMAV	Belice	SLC	SLC_no_Andino	SI	SI	NO	rojo	93.0	109.7	2.6	0.3	2.4	3.1	24.1	17.9	1.1	1.5	120.5	132.2

BGV8230	COMAV	Honduras	SLC	SLC_no_Andino	SI	SI	NO	rojo	43.0	101.3	0.2	5.2	1.2	2.6	7.6	17.0	0.5	2.3	52.3	123.2
CR001	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	47.8	32.9	1.1	3.5	1.4	1.1	49.7	30.8	2.0	1.5	100.9	66.2
CR002	NA	NA	NA	NA	NO	NO	SI	NA	112.2	91.7	7.0	3.6	3.0	4.9	20.8	7.0	1.4	1.8	136.0	105.5
CR028	INRA	NA	Mixto	Mixto	SI	SI	NO	rojo	47.9	37.5	2.4	4.7	1.4	1.2	14.7	8.4	1.2	1.0	65.3	48.1
CR049	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	61.4	46.2	15.1	13.2	6.0	7.0	29.3	19.3	2.4	1.6	99.1	72.8
CR056	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	NA	86.8	99.6	2.0	1.7	2.3	2.3	14.0	7.2	0.3	0.9	103.4	110.0
CR058	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	49.4	42.6	2.4	1.5	1.6	1.2	4.5	2.3	0.2	0.7	55.7	46.8
CR061	TGRC	Perú	SP	SP_Perú	SI	SI	NO	rojo	35.9	33.1	6.5	7.2	1.7	2.9	5.1	9.9	1.3	0.4	44.1	46.3
CR064	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	NA	49.4	60.9	2.1	2.4	0.8	0.7	67.5	45.5	5.5	1.7	123.3	108.7
CR073	VIR	Checoslovaquia	Mixto	Mixto	SI	SI	SI	rojo	51.4	55.2	11.2	25.1	4.1	3.3	8.7	10.1	1.7	5.3	66.0	73.9
CR075	VIR	Checoslovaquia	SLC	SLC_no_Andino	SI	SI	SI	naranja	126.2	166.1	0.8	2.0	4.9	5.4	11.3	7.6	1.3	1.8	143.7	180.9
CR076	VIR	Alemania	SLC	SLC_no_Andino	SI	SI	SI	amarillo	108.2	158.6	2.3	2.1	6.9	6.5	32.4	30.3	5.9	5.7	153.4	201.2
CR077	VIR	Perú	Mixto	Mixto	SI	SI	SI	rojo	47.4	51.2	19.5	14.2	2.6	5.4	23.0	24.6	4.3	2.6	77.3	83.8
CR078	VIR	Argentina	SLC	SLC_no_Andino	SI	SI	SI	rojo	69.1	47.4	24.9	14.8	4.2	5.9	28.3	21.4	4.8	3.2	106.3	77.9
CR079	VIR	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	51.3	73.7	15.4	14.2	4.9	1.2	17.9	19.4	3.1	5.9	77.2	100.2
CR081	VIR	Mexico	SLC	SLC_no_Andino	SI	SI	SI	rojo	76.0	79.1	1.4	2.3	2.0	2.1	24.9	17.4	1.3	1.0	104.2	99.6
CR082	VIR	Guatemala	SLC	SLC_no_Andino	SI	SI	SI	rojo	62.6	64.3	0.3	1.4	1.8	1.8	24.0	13.4	1.2	0.9	89.6	80.3
CR085	VIR	Guatemala	Mixto	Mixto	SI	SI	SI	rojo	57.7	78.8	24.3	16.0	1.4	4.0	17.7	17.9	6.6	2.6	83.5	103.2
CR086	VIR	Guatemala	SLC	SLC_no_Andino	SI	SI	SI	rojo	32.5	37.4	0.9	4.3	0.7	0.8	44.2	45.7	2.7	3.4	80.1	87.3
CR088	VIR	Tunes	SLC	SLC_no_Andino	SI	SI	SI	amarillo	161.0	92.4	1.6	1.4	9.3	5.3	45.5	36.2	4.4	4.5	220.3	138.4
CR089	VIR	Dinamarca	SLC	SLC_no_Andino	SI	SI	SI	amarillo	91.7	103.2	1.6	1.6	4.9	4.3	38.9	40.3	2.8	4.6	138.3	152.4
CR092	VIR	Rusia	SLC	SLC_no_Andino	SI	SI	SI	rojo	76.0	88.4	1.0	0.7	1.4	2.7	11.3	10.4	0.8	1.3	89.6	102.8
CR093	VIR	NA	SLC	SLC_no_Andino	SI	NO	NO	rojo	60.8	68.1	18.8	13.6	4.2	7.8	18.1	20.3	4.4	0.7	87.5	96.9
CR095	VIR	Perú	Mixto	Mixto	SI	SI	SI	rojo	62.6	53.6	16.6	13.8	8.1	7.8	39.8	41.0	1.2	1.1	111.7	103.5
CR097	VIR	NA	SP	conflicting	SI	SI	SI	rojo	45.8	56.1	16.5	12.5	6.2	6.1	17.1	13.0	2.4	0.5	71.5	75.7
CR098	VIR	Perú	SLC	SLC_Perú	SI	NO	NO	rosado	88.5	102.3	2.4	1.5	1.9	2.0	30.7	34.8	0.9	2.0	122.1	141.1
CR099	VIR	Perú	SLC	SLC_Perú	SI	SI	SI	rojo	113.2	97.6	1.7	2.6	4.1	3.2	40.9	24.7	0.1	2.0	158.3	127.5
CR100	VIR	Cuba	SLC	SLC_no_Andino	SI	SI	SI	rosado	64.7	74.1	1.1	0.9	8.6	2.0	13.9	9.6	0.4	0.9	87.5	86.6
CR101	VIR	NA	SLC	SLC_no_Andino	SI	NO	NO	rojo	93.9	98.7	2.6	1.2	2.3	2.1	17.5	18.5	1.1	1.3	114.8	120.6
CR103	TGRC	Honduras	SLC	SLC_no_Andino	SI	SI	SI	rojo	72.9	82.9	2.9	1.5	7.4	1.5	16.2	11.4	1.3	1.8	97.8	97.6
CR104	TGRC	Santa Cruz	SLC	SLC_Ecuador	SI	SI	SI	rojo	75.4	74.4	15.1	15.3	8.6	2.9	14.8	13.2	0.7	4.4	99.5	95.0
CR105	TGRC	Hawai	SLC	SLC_no_Andino	SI	SI	SI	rojo	61.8	67.3	2.2	1.6	1.4	1.3	14.7	11.8	0.2	1.3	78.1	81.7
CR106	TGRC	Guatemala	SLC	SLC_no_Andino	SI	NO	NO	rosado	53.2	79.2	0.6	1.1	1.0	1.4	6.1	15.3	0.5	1.1	60.8	97.1
CR107	TGRC	Ecuador	SLC	SLC_Perú	SI	SI	SI	rojo	95.3	87.0	0.6	2.2	1.8	1.3	9.0	12.5	0.5	1.8	106.6	102.6

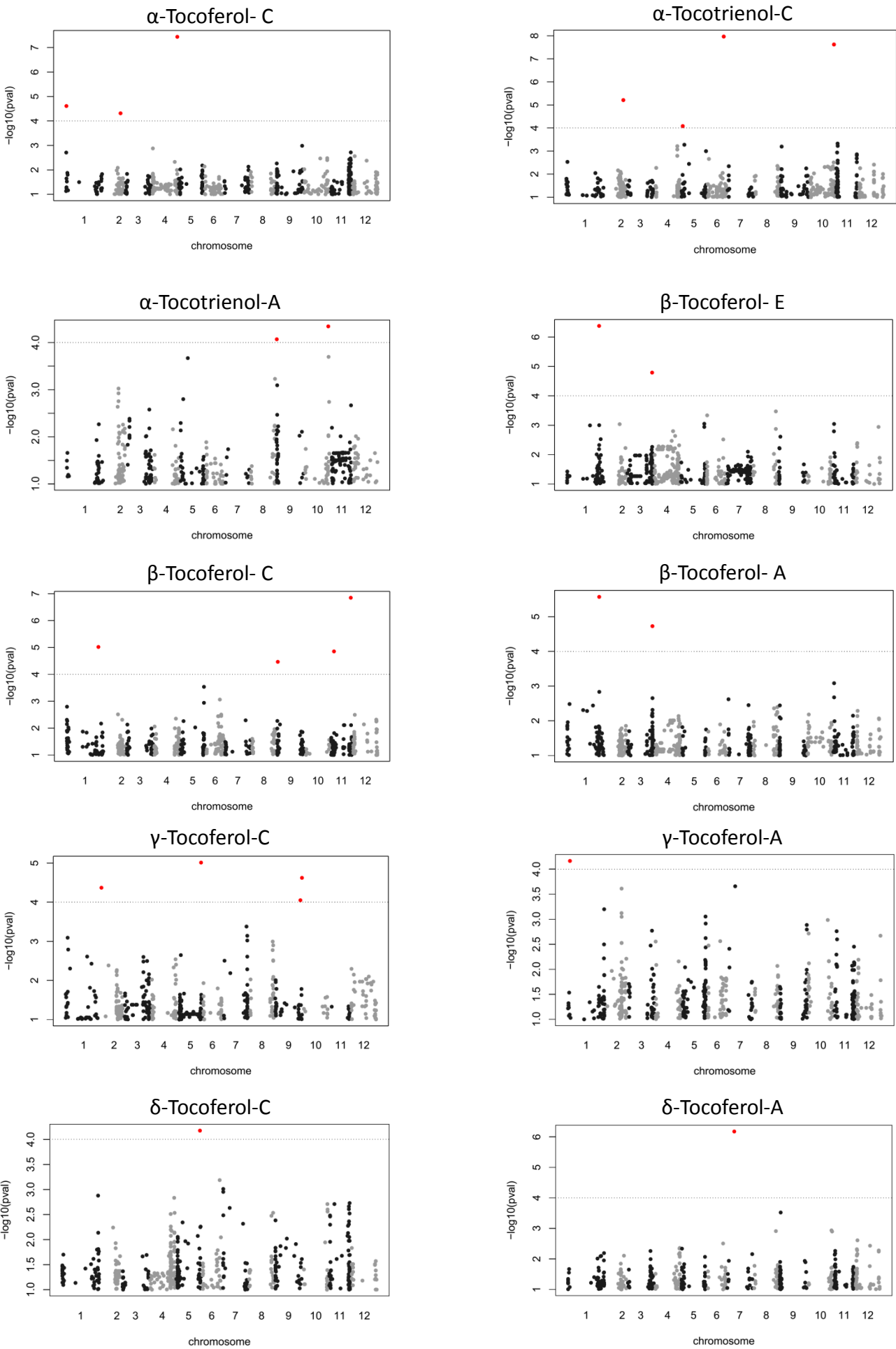
CR108	TGRC	Ecuador	Mixto	Mixto	SI	SI	NO	rojo	44.7	53.1	14.0	11.6	4.3	6.0	27.2	12.5	2.8	0.5	78.9	72.1
CR110	TGRC	Perú	SLC	SLC_Perú	SI	SI	NO	rojo	59.3	67.0	13.5	19.4	3.6	3.1	19.7	9.8	4.3	4.5	87.0	84.4
CR111	NA	NA	NA	NA	NO	NO	NO	NA	75.0	54.1	0.6	5.1	1.3	0.9	30.3	16.6	1.3	1.0	107.9	72.7
CR113	NA	NA	NA	NA	NO	NO	NO	NA	72.2	58.6	0.8	4.6	2.1	2.1	21.6	12.0	1.7	0.7	97.6	73.4
CR114	TGRC	Perú	SLC	SLC_Perú	SI	NO	NO	rojo	59.5	38.5	2.1	6.5	1.6	0.9	15.6	13.7	1.2	1.0	77.9	54.1
CR115	TGRC	Ecuador	SLC	SLC_Ecuador	SI	SI	SI	rojo	56.4	66.4	11.2	16.7	5.1	4.9	12.3	11.3	2.1	3.1	75.8	85.6
CR117	TGRC	Perú	SLC	SLC_Perú	SI	SI	NO	rojo	49.9	56.0	1.7	1.8	1.4	1.7	37.6	38.6	2.1	1.9	91.1	98.2
CR118	TGRC	Ecuador	SLC	SLC_Ecuador	SI	SI	NO	rosado	88.8	84.9	0.8	0.9	2.4	2.9	10.8	9.7	0.5	1.4	102.6	98.8
CR119	NA	NA	NA	NA	NO	NO	NO	NA	195.3	63.9	3.2	5.2	3.5	2.4	35.0	7.5	1.5	1.1	235.3	74.8
CR120	TGRC	Ecuador	SP	SP_Ecuador	SI	SI	SI	rojo	52.8	85.9	14.7	17.2	4.0	9.9	17.1	16.1	2.5	0.7	76.4	112.7
CR122	TGRC	Mexico	SP	SLC_SP_Perú	SI	NO	NO	rojo	33.7	30.3	14.2	13.9	2.5	0.4	27.3	24.2	7.0	4.8	70.5	59.7
CR123	TGRC	Filipinas	SLC	SLC_no_Andino	SI	SI	NO	rojo	130.7	134.8	2.6	0.8	4.2	4.1	29.0	25.7	0.9	1.7	164.9	166.3
CR124	TGRC	Honduras	SLC	SLC_no_Andino	SI	SI	NO	rojo	123.9	127.7	0.8	1.5	3.8	4.0	8.8	18.6	0.8	1.6	137.3	151.9
CR125	TGRC	Malasia	SLC	SLC_no_Andino	SI	NO	NO	rojo	109.0	83.2	3.1	2.5	3.3	2.6	20.2	14.6	1.4	1.3	133.9	101.6
CR130	TGRC	Bolivia	SLC	SLC_Perú	SI	NO	NO	amarillo	160.1	160.1	2.1	5.5	9.1	8.7	34.5	23.1	5.1	6.9	208.8	198.8
CR141	TGRC	Perú	SLC	SLC_Perú	SI	NO	NO	rojo	70.8	80.6	16.4	16.3	4.2	7.1	17.8	17.1	4.6	1.9	96.7	106.7
CR142	TGRC	Malasia	SLC	SLC_no_Andino	SI	SI	SI	rosado	98.5	111.5	6.1	2.4	3.6	4.4	17.2	15.3	3.0	1.3	122.4	132.5
CR144	TGRC	Costa Rica	SLC	SLC_no_Andino	SI	SI	SI	rojo	72.2	67.8	2.6	1.7	1.7	1.4	38.3	31.3	2.3	2.0	114.5	102.5
CR145	TGRC	Brasil	SLC	SLC_no_Andino	SI	SI	SI	rojo	121.1	73.4	1.7	1.6	2.7	3.7	10.2	15.4	0.5	1.4	134.4	83.6
CR147	TGRC	Zambia	SLC	SLC_no_Andino	SI	SI	SI	rojo	108.0	91.8	0.7	10.9	2.4	3.9	22.4	22.3	1.2	3.0	134.0	121.0
CR148	TGRC	Brazil	SLC	SLC_Perú	SI	SI	SI	rojo	44.6	61.6	15.8	13.9	5.2	3.4	19.4	21.4	3.1	4.9	71.6	91.2
CR149	TGRC	Ecuador	SLC	SLC_Ecuador	SI	SI	SI	rojo	77.4	79.6	1.8	3.4	1.2	1.7	20.4	16.8	0.9	1.2	99.8	99.3
CR152	TGRC	Perú	SLC	SLC_Perú	SI	SI	SI	rojo	52.1	53.0	17.6	18.3	10.2	3.8	17.0	12.8	0.9	4.2	80.3	73.8
CR153	TGRC	Perú	SLC	SLC_Perú	SI	SI	SI	rojo	50.5	50.4	12.7	16.8	7.5	3.5	11.3	13.5	0.6	4.9	70.0	72.3
CR154	TGRC	Indonesia	SLC	SLC_no_Andino	SI	SI	SI	rojo	80.1	193.7	2.3	2.7	2.2	5.6	25.5	39.3	2.2	3.5	110.0	242.0
CR155	TGRC	Brasil	SLC	SLC_Perú	SI	SI	SI	rojo	126.6	163.9	1.0	0.2	5.6	5.7	12.7	4.4	1.5	0.7	146.5	174.7
CR156	TGRC	Perú	Mixto	Mixto	SI	SI	SI	rojo	46.9	52.2	14.9	19.9	4.4	1.0	37.5	35.6	2.8	6.7	91.6	95.4
CR157	TGRC	Perú	SLC	SLC_Perú	SI	SI	SI	rojo	86.2	86.1	1.2	1.5	1.9	2.1	16.3	6.7	0.8	0.4	105.1	95.4
CR158	TGRC	Perú	SP	SLC_SP_Perú	SI	SI	SI	rojo	26.8	39.0	13.0	15.0	3.5	2.6	24.9	23.1	2.2	4.2	57.3	68.9
CR161	TGRC	India	Mixto	Mixto	SI	SI	SI	rojo	60.2	52.2	7.8	14.5	4.8	5.3	10.6	15.3	0.7	2.8	76.3	75.5
CR166	TGRC	Ecuador	SP	SP_Perú	SI	SI	NO	rojo	24.3	36.9	10.2	10.4	4.5	8.9	20.0	25.4	0.4	2.9	49.2	74.1
CR179	TGRC	Perú	SP	SP_Perú	SI	SI	NO	rojo	69.8	71.3	3.5	5.9	0.5	2.2	9.4	4.8	0.6	1.5	80.3	79.8
CR181	TGRC	Perú	SP	SP_Perú	SI	SI	NO	rojo	23.9	21.6	8.7	13.0	3.5	0.2	26.1	23.9	0.7	4.6	54.3	50.3
CR187	NA	NA	NA	NA	NO	NO	NO	NA	47.7	82.6	0.4	3.5	0.6	1.3	41.8	37.2	1.5	1.8	91.5	122.9

CR199	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rosado	87.4	83.2	1.3	0.7	2.5	2.2	14.7	5.6	1.0	1.0	105.5	92.0
CR202	CGN	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	103.8	89.8	1.0	3.5	2.4	2.2	26.1	19.8	1.1	1.5	133.4	113.3
CR203	TGRC	Perú	SP	SP_Perú	SI	SI	SI	rojo	76.2	23.6	21.0	15.0	3.5	6.5	29.9	17.6	7.8	0.6	117.4	48.3
CR205	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	26.8	31.7	18.2	21.5	3.0	4.6	21.2	21.1	5.3	4.9	55.5	62.2
CR236	USDA	Perú	SLC	SLC_Perú	SI	SI	SI	rojo	76.5	75.7	1.2	1.4	1.8	1.9	15.7	20.3	1.2	2.0	95.1	99.9
CR238	USDA	Colombia	SLC	SLC_no_Andino	SI	SI	NO	rojo	49.7	55.4	11.7	10.5	6.1	2.5	10.6	9.7	0.5	3.1	66.9	70.8
CR240	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	46.7	49.5	16.5	13.5	3.7	4.1	8.1	6.2	3.3	1.7	61.7	61.4
CR241	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	59.7	61.4	10.3	15.1	3.5	4.7	22.4	10.4	2.3	2.8	87.8	79.4
CR242	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	460.5	303.6	6.0	2.7	12.0	10.4	25.5	7.5	1.5	0.9	499.4	322.4
CR245	NA	NA	NA	NA	NO	NO	NO	NA	59.2	48.0	3.1	4.4	1.7	19.8	38.1	34.7	1.9	2.1	100.9	104.6
CR246	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	27.2	31.2	17.0	10.8	0.6	3.3	23.5	24.3	5.7	2.6	56.9	61.3
CR248	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	55.0	61.1	12.6	11.5	4.5	2.6	9.4	9.6	2.6	2.2	71.5	75.5
CR250	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	95.4	50.2	3.0	4.5	3.0	1.5	31.2	18.2	1.9	2.2	131.6	72.1
CR253	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	1.6	67.6	0.0	4.6	0.0	1.9	0.9	37.9	0.0	2.6	2.6	110.1
CR263	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	56.7	66.7	14.2	14.1	5.7	7.4	14.2	8.4	2.1	0.5	78.7	83.0
CR264	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	83.8	73.6	2.4	4.9	3.9	4.0	22.7	19.6	2.3	2.3	112.7	99.5
CR271	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	54.1	34.4	1.3	4.9	2.2	1.3	11.9	7.2	0.5	1.2	68.7	44.2
CR274	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	69.9	72.1	3.0	2.9	1.8	2.0	22.3	17.7	1.9	1.6	95.2	92.7
CR275	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	amarillo	123.6	65.0	2.2	6.3	4.5	2.7	11.7	8.0	1.3	2.7	141.0	78.4
CR276	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	91.2	57.0	0.5	3.7	2.5	2.4	12.8	9.4	0.8	1.0	107.4	69.7
CR282	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	45.9	36.9	21.4	14.8	2.9	1.2	28.4	20.9	7.8	6.5	85.0	65.4
CR283	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	62.9	37.4	0.8	4.6	1.8	1.5	36.1	22.8	1.7	1.7	102.5	63.4
CR286	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	54.9	62.3	1.8	3.7	1.4	1.3	20.9	1.9	1.2	0.3	78.4	65.9
CR287	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	97.6	73.8	1.9	4.5	2.5	2.0	28.9	17.3	1.8	1.3	130.8	94.3
CR288	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	47.1	71.9	2.4	1.6	1.1	1.7	20.4	17.1	0.6	2.1	69.3	92.7
CR289	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	105.3	72.6	1.2	4.3	2.5	1.7	8.9	5.5	0.8	1.0	117.5	80.8
CR290	INRA	NA	Mixto	Mixto	SI	SI	SI	amarillo	56.8	78.3	19.7	13.8	6.6	7.0	27.9	21.6	5.7	4.4	96.9	111.3
CR291	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rosado	72.8	37.3	2.3	5.8	3.2	2.2	9.9	3.2	0.9	0.8	86.8	43.5
CR292	INRA	NA	Mixto	Mixto	SI	SI	SI	rojo	54.5	38.3	12.0	16.1	6.5	3.4	20.0	16.5	2.1	4.3	83.0	62.6
CR293	INRA	NA	SLC	SLC_no_Andino	SI	SI	SI	rojo	71.2	91.5	2.8	15.7	2.7	3.6	18.7	14.7	1.6	3.9	94.2	113.7
CR354	NA	NA	NA	NA	NO	NO	SI	NA	69.6	74.5	8.4	1.3	12.6	2.6	37.5	15.6	1.9	2.8	121.6	95.5
T102068	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	115.5	83.6	1.1	3.9	3.5	2.6	7.2	3.2	0.5	1.3	126.6	89.7
T102950	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rosado	76.5	46.3	1.9	4.5	2.3	1.4	15.1	9.1	0.8	0.9	94.7	57.8
T102957	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	73.3	50.9	0.7	4.0	1.4	1.1	29.7	16.8	1.0	1.1	105.5	69.9

T102958	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	67.9	57.3	0.8	2.5	1.5	1.5	32.8	20.2	1.3	1.1	103.5	80.1
T102959	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	98.0	60.6	1.9	4.8	1.9	1.4	25.2	6.5	1.3	1.0	126.4	69.5
T102961	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	100.7	88.3	2.8	2.6	1.8	1.7	9.5	8.8	1.6	0.7	112.9	99.6
T102962	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	82.5	52.4	3.2	2.8	1.7	1.2	24.1	10.9	0.4	0.7	108.6	65.1
T102963	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rosado	106.1	82.2	1.7	2.8	3.0	2.1	16.4	8.0	0.8	1.1	126.3	93.4
T102964	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	amarillo	145.6	103.6	1.5	8.0	3.8	3.8	11.7	4.4	0.9	3.1	162.0	114.9
T102965	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	69.8	67.6	2.6	4.0	1.7	1.6	20.6	7.0	0.4	0.8	92.5	77.0
T102966	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	55.1	70.1	3.9	2.6	1.2	1.5	13.7	8.3	0.4	0.6	70.4	80.5
T102967	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	108.9	91.3	2.8	3.1	2.7	1.9	32.3	15.6	2.1	1.2	145.9	110.0
T102969	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	61.3	107.0	2.4	3.1	1.7	2.9	12.8	7.9	0.3	0.7	76.2	118.4
T102970	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	56.1	88.8	3.8	1.5	1.3	2.2	11.5	14.8	0.2	1.1	69.1	106.9
T102971	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	87.5	119.5	1.2	3.3	4.1	5.1	32.0	28.6	1.2	2.6	124.8	155.8
T102972	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	92.3	93.4	2.0	2.9	4.3	4.6	35.6	20.3	3.8	3.4	136.0	121.6
T102973	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	NA	104.9	93.7	1.4	2.2	1.9	2.1	10.1	9.4	0.6	1.1	117.5	106.3
T102974	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	60.4	78.9	1.5	1.4	1.0	1.4	6.4	13.1	0.6	0.8	68.3	94.2
T102975	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	101.6	109.0	2.1	7.0	1.4	2.7	17.0	6.9	0.0	1.5	120.0	120.2
T102976	NA	NA	NA	NA	NO	NO	NO	NA	69.0	93.2	2.0	2.5	3.0	1.7	13.2	4.3	0.2	1.1	85.3	100.3
T102977	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	99.4	144.8	1.6	1.8	1.9	2.1	6.1	5.8	0.0	0.6	107.5	153.3
T102979	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	54.0	69.9	9.2	9.2	2.5	2.4	12.8	10.3	2.6	3.1	71.8	85.7
T102980	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	rojo	79.5	109.5	1.4	0.3	1.4	2.2	9.8	7.9	0.9	0.9	90.9	120.5
T102981	NA	NA	NA	NA	NO	NO	NO	NA	64.1	45.3	5.1	3.4	2.3	1.7	8.6	4.1	0.6	1.1	75.5	52.2
T102982	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	amarillo	144.1	134.8	27.1	0.7	4.2	3.9	9.6	15.6	1.2	1.4	158.4	155.8
T102983	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	amarillo	225.3	158.1	2.1	5.3	5.1	4.8	21.1	21.0	0.7	3.6	252.2	187.4
T102984	MALAGA	Mexico	SLC	SLC_no_Andino	SI	NO	NO	amarillo	45.9	120.4	9.5	8.4	1.2	5.7	9.2	18.3	4.5	2.4	60.8	146.7

Tabla S2. Materiales GWAS: banco de germoplasma; origen geográfico; grupo genético; sub-grupo genético, referencias, color de fruto; contenidos de tococromanos en frutos. ⁽¹⁾ Identificador de la accesión de la población de GWAS referido según el banco de germoplasma ⁽²⁾ : TGRC para 'Tomato Genetics Resource Center' de la Universidad de Davis (USA); INRA para 'National Agricultural Research Institute of Avignon' (Francia); MALAGA para el 'department of Molecular Biology and Biochemistry' Universidad de Malaga (España); VIR referido al 'Vavilov Institute of Plant Genetic Resources of St. Petersburg' (Rusia); COMAV para 'Institute for Conservation and Improvement of Valencian agrobiodiversity of Valencia' (España), USDA referido al 'United States Department of Agriculture' (USA) y CGN referido al 'Centre Genetic Resources of Wageningen' (Holanda). ^{(3),(4)} Grupo y Sub-grupo genético o clasificación genética propuesta por Blanca y col.(2015). ^{(6),(7),(8)} Referencias de otros trabajos donde fueron utilizados estos materiales previamente, SI y NO define si está o no presente respectivamente entre los 146 materiales estudiados en esta tesis. ⁽⁹⁾ Media de los contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total medidos por HPLC expresados como [$\mu\text{g/gPS}$], μg de la forma de tococromanol por gramo de tejido seco, calculada a partir de 2 a 3 réplicas por material y condición de riego: **C** refiere a los materiales cultivados en condiciones control de riego y **E** a condiciones de estrés hídrico. ⁽¹⁰⁾ NA refiere dato desconocido o faltante.

Figura S1. Manthattan plots para las 32 asociaciones determinadas en el GWAS2



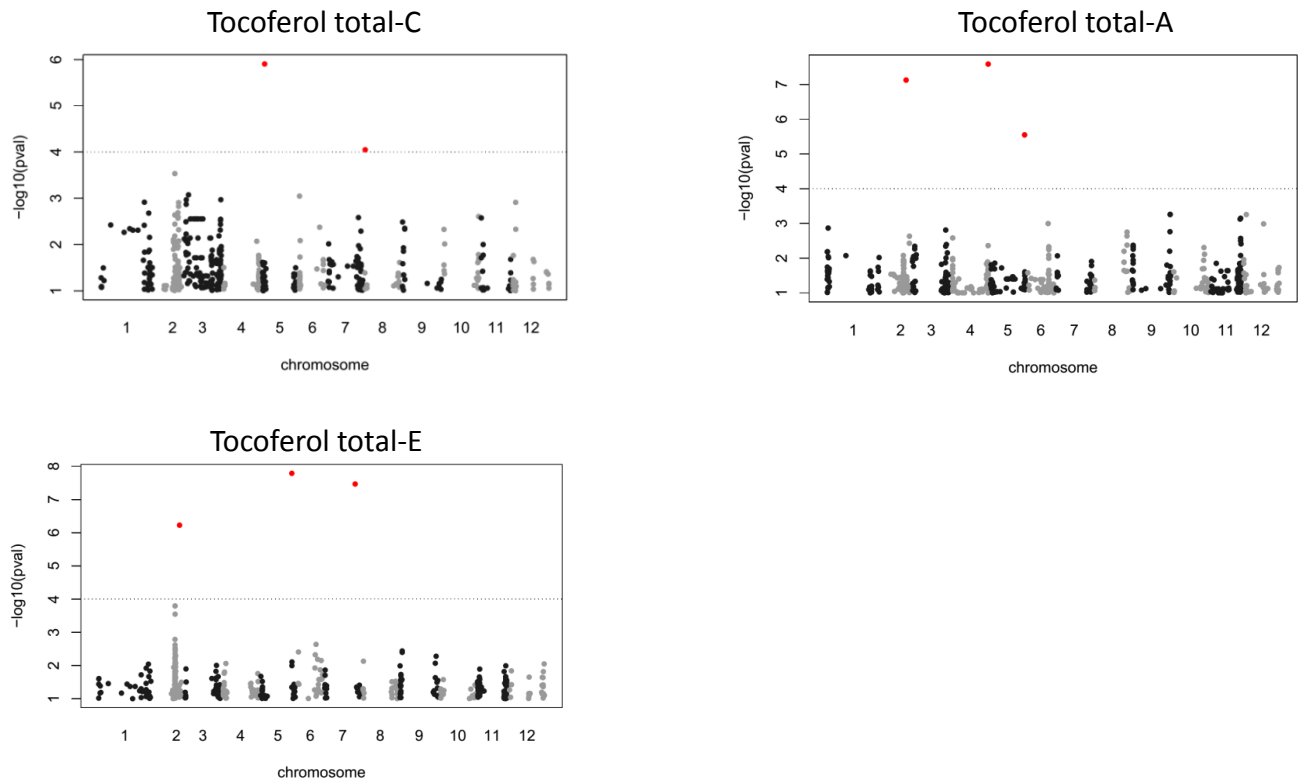
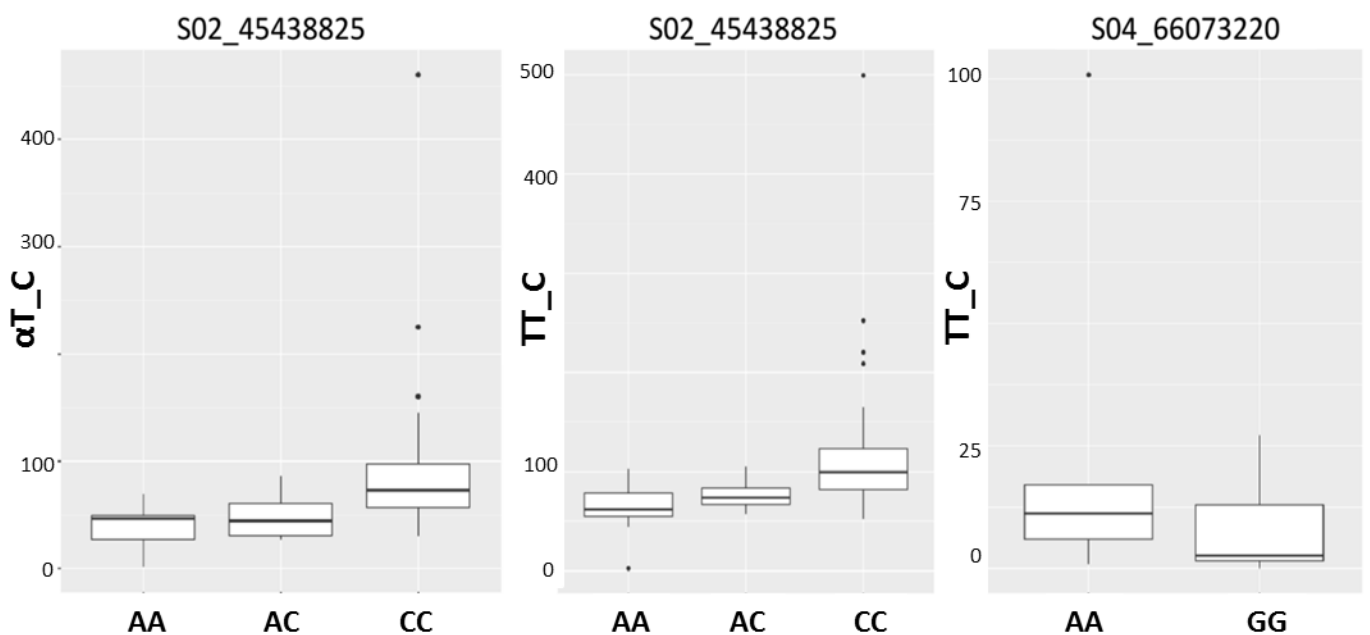
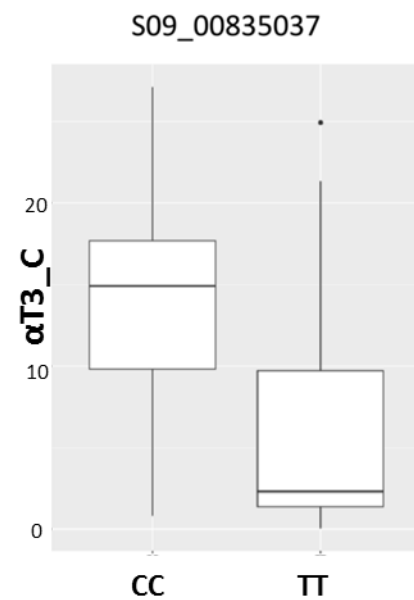
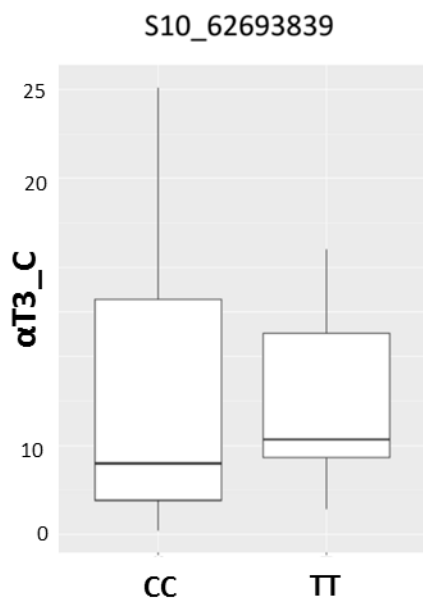
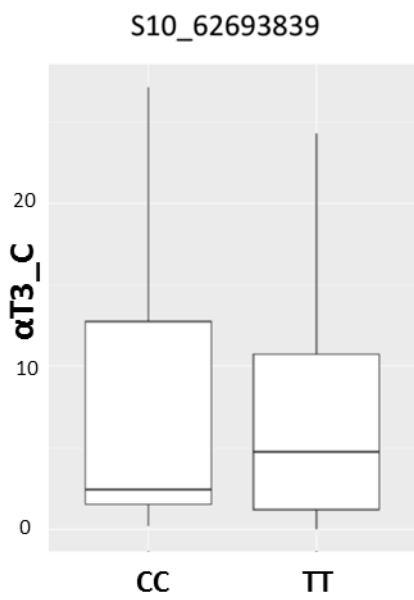
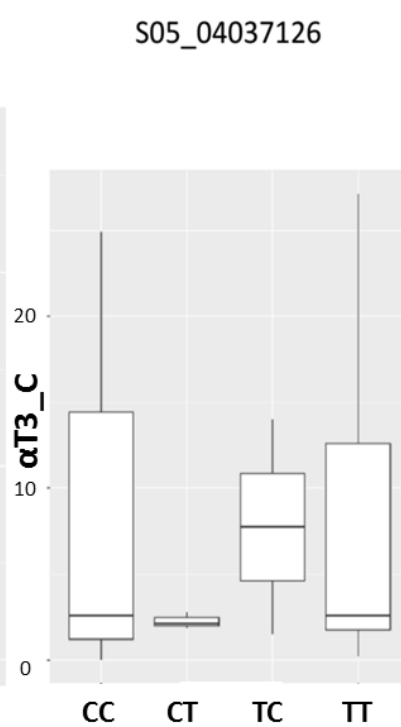
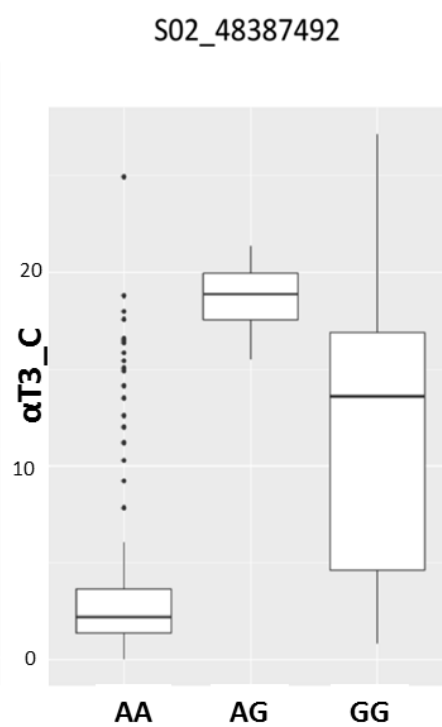
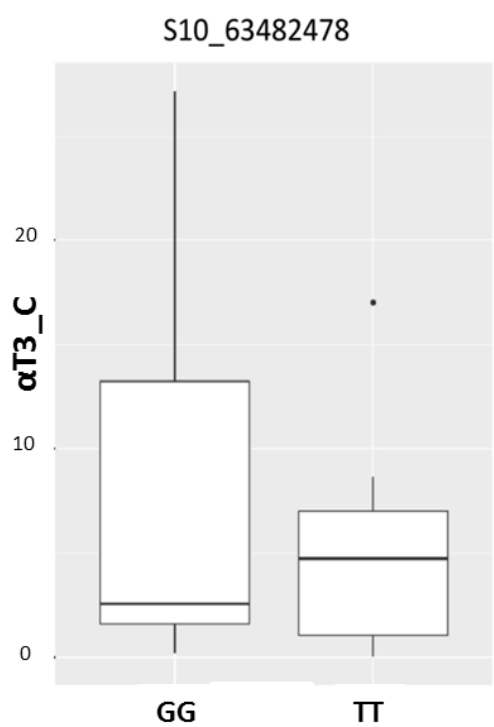
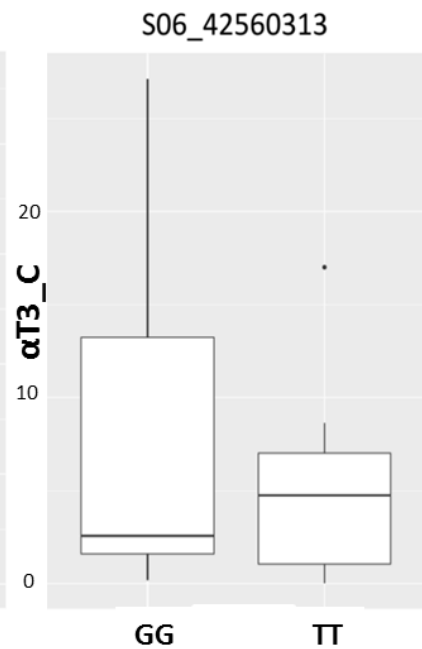
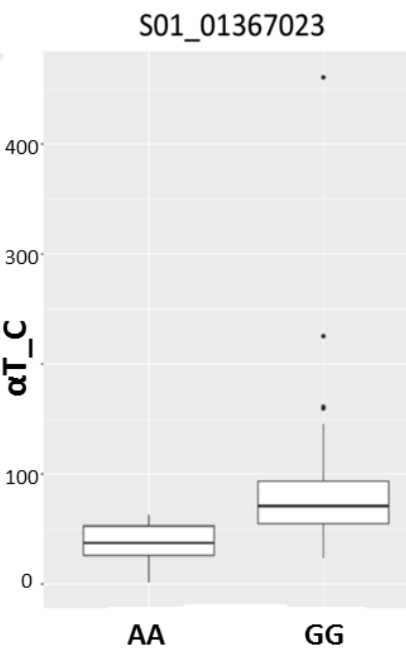
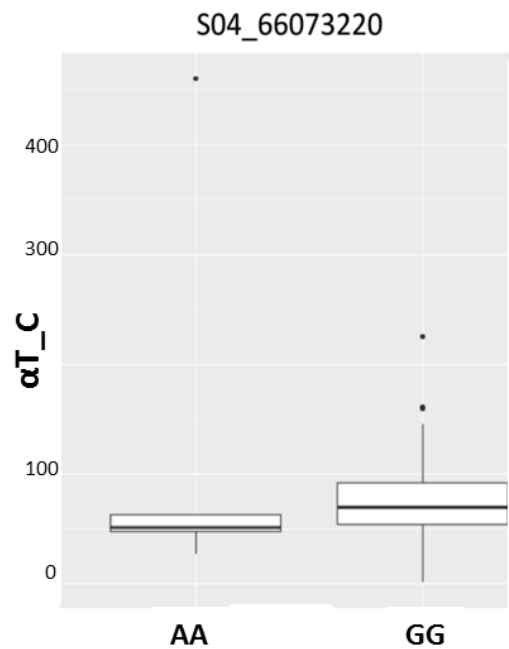


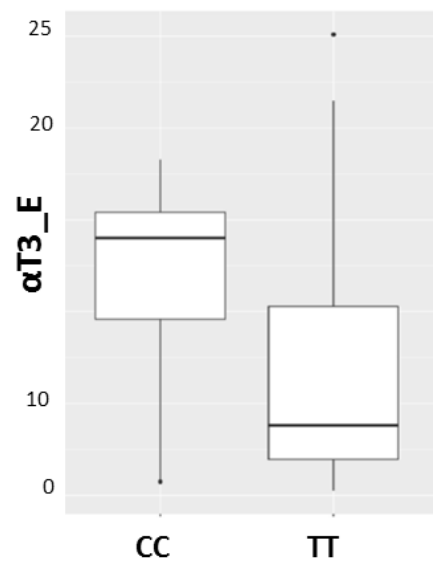
Figura S1. *Manhattan plot* para las 32 asociaciones identificadas asociadas a tococromanoles en el GWAS2. Los 32 marcadores significativos asociados a cada isóforma tococromanol determinados a partir del GWAS 2 para el conjunto dentro de los 146 materiales evaluados, se encuentran en 13 *manhattan plots* titulados en la parte superior de cada uno de los graficos de ellos, para ambos ambientes C (control) y E (estrés). Se puede observar la línea de corte utilizada en el análisis GWAS 2 con un pValor de 10^{-4} . En el eje de abscisas se puede observar la ubicación de los 6.100 SNPs a lo largo de los 12 cromosomas del genoma.

Figura S2. *Box plots* para las 32 asociaciones genotipo-fenotipo determinadas a partir del análisis de GWAS2

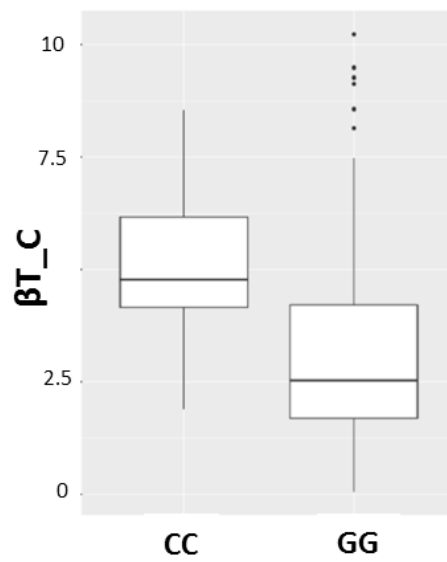




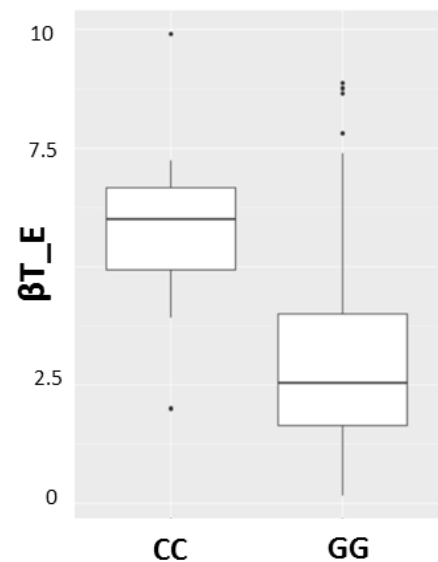
S09_00835037



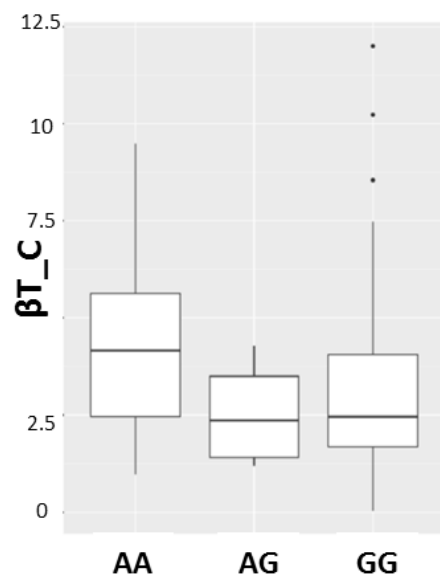
S01_85997795



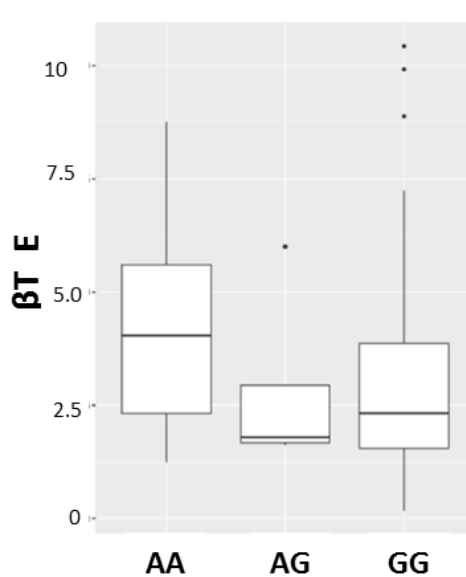
S01_85997795



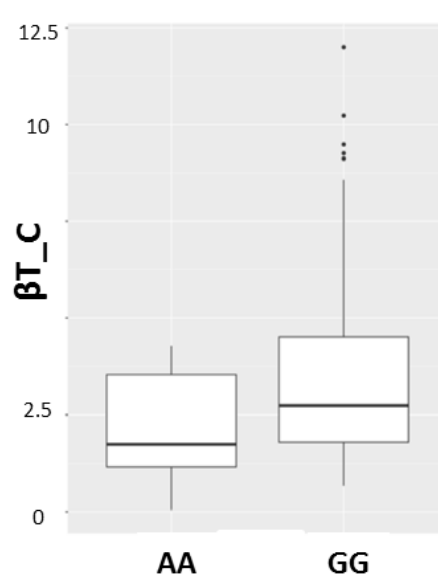
S03_66764638



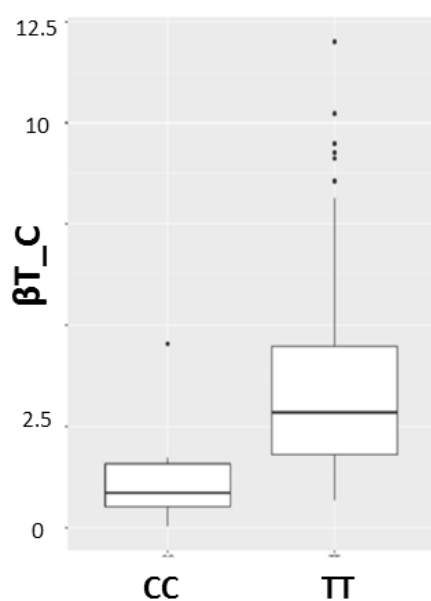
S03_66764638



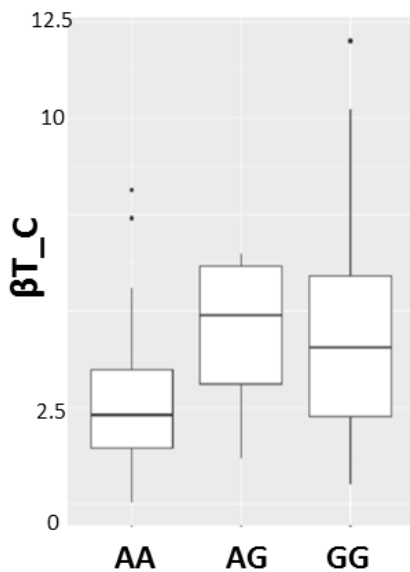
S11_52844288



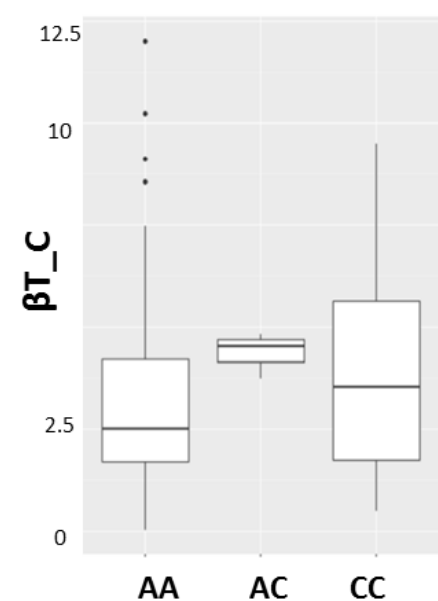
S09_01517398

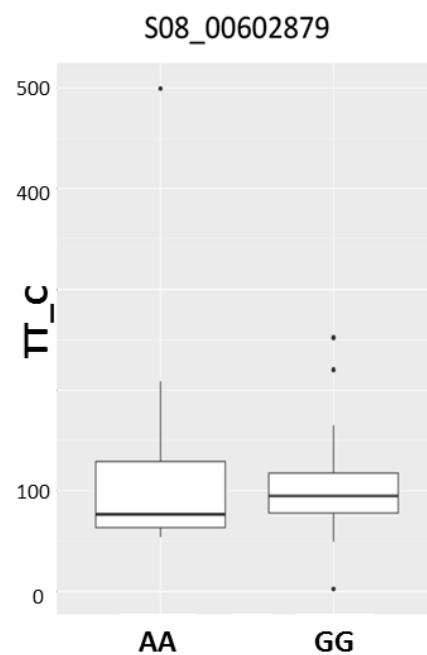
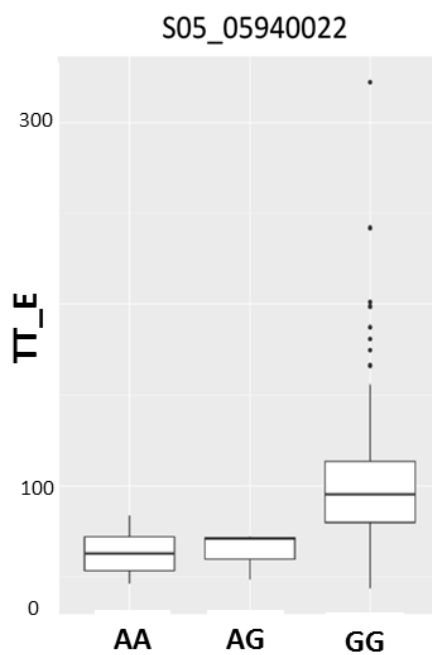
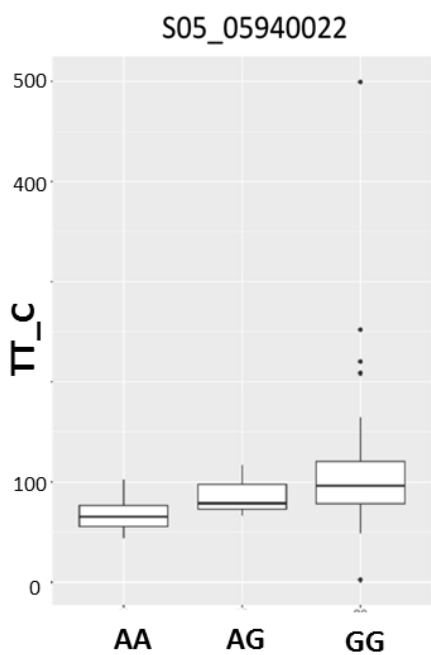
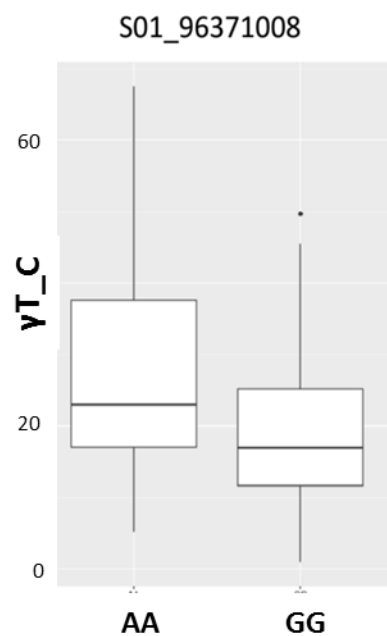
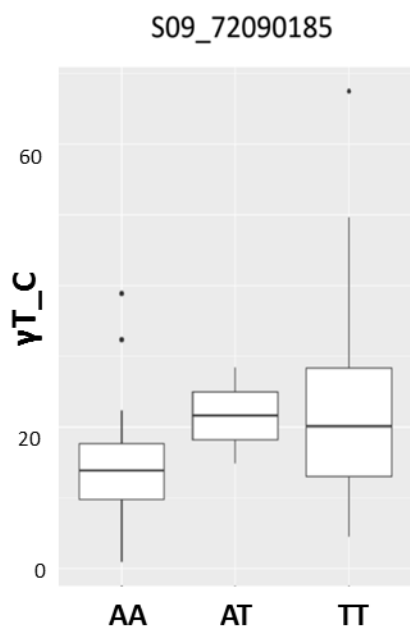
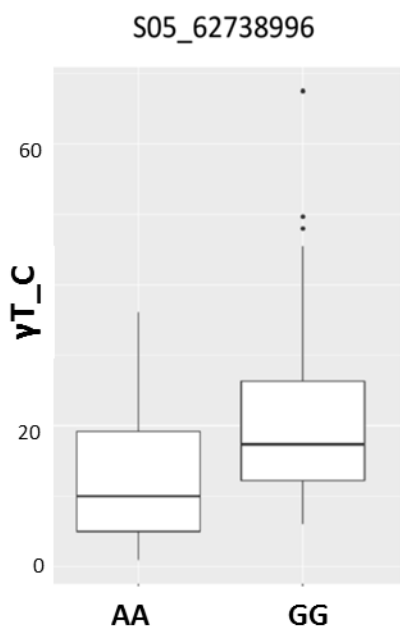
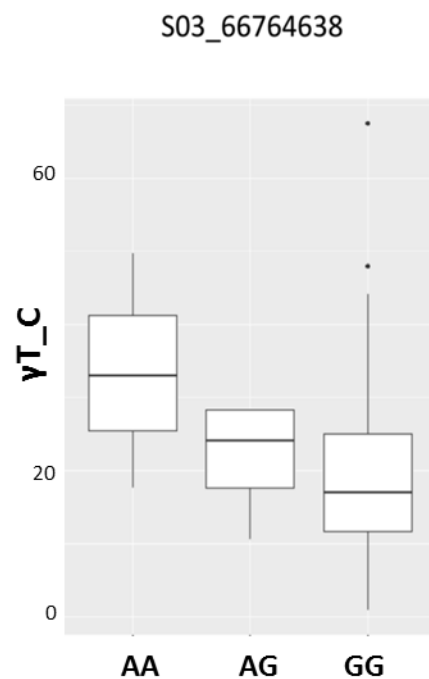
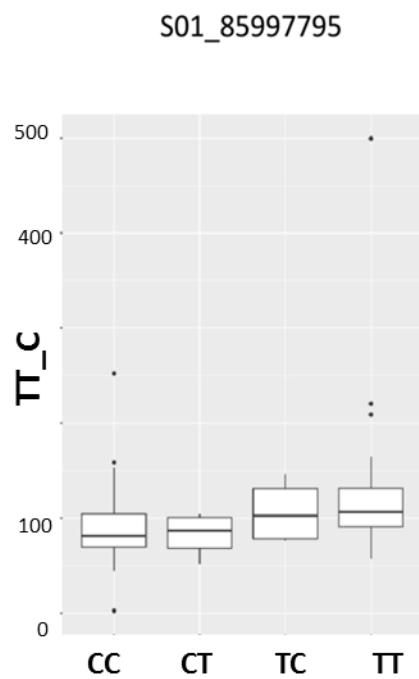
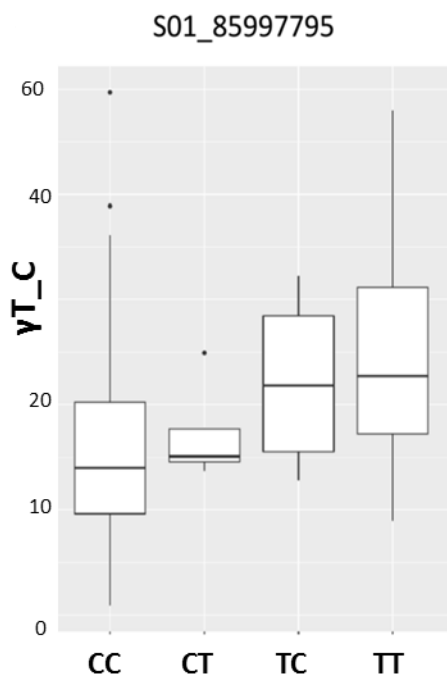


S01_85349943



S11_07669410





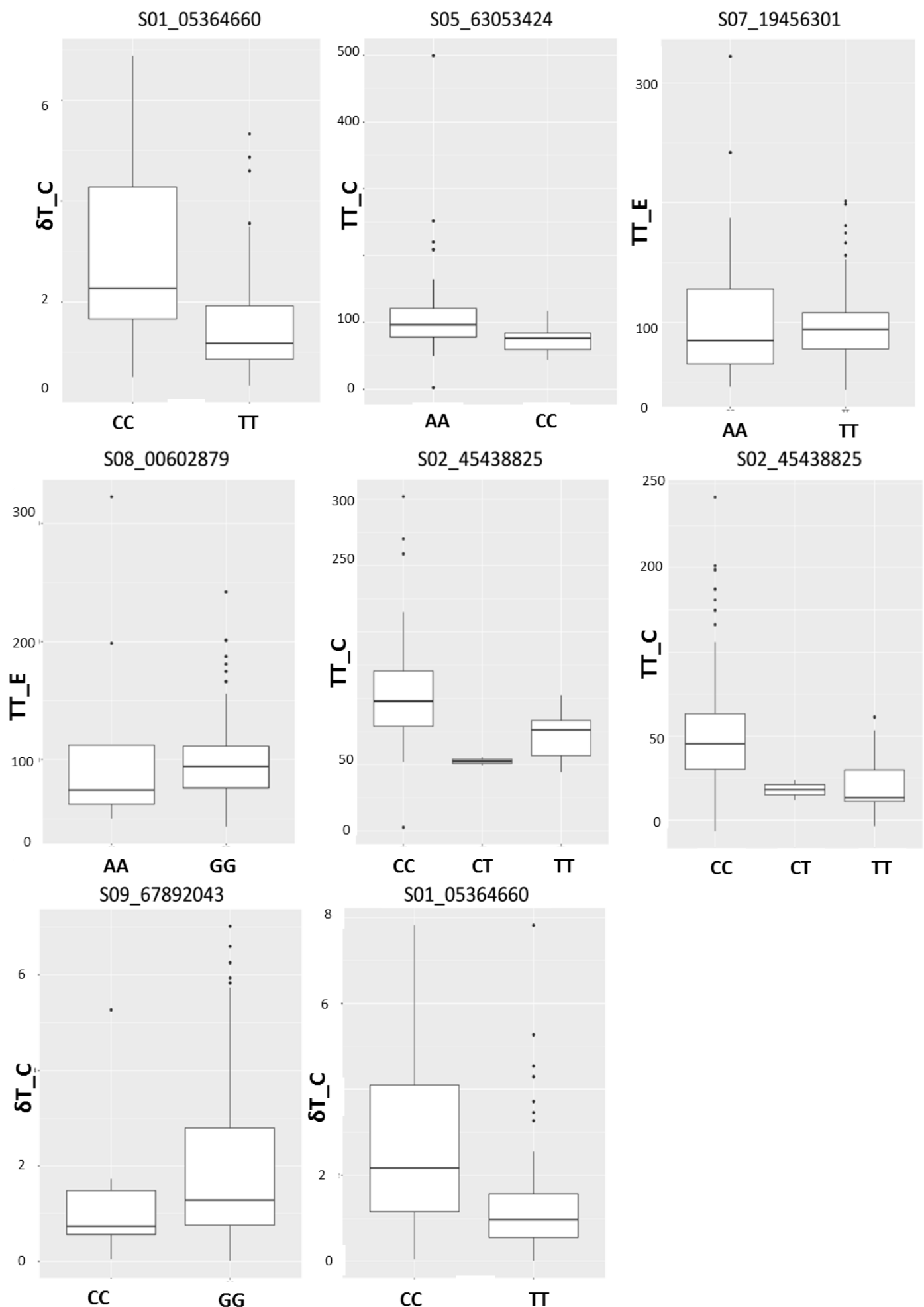


Figura S2. Box plots para las 32 asociaciones genotipo-fenotipo determinadas a partir del análisis de GWAS2. Graficos de boxplots para las distribuciones de las isformas de αT (alfa tocoferol) αT₃ (alfa tocotrienol); βT (beta tocoferol); γ-T (gama tocoferol); δT (delta tocoferol); TT (tocoferol total) de frutos de las 146 entradas del GWAS en condiciones control (C) y de estrés (E) según su haplotipo. En el eje de ordenadas se encuentra la escala de unidades [ug/gPS] y en el de abscisas el polimorfismo (SNP) asociado.

Tabla S3. Datos fenotípicos, contenidos de tococromanoles, de la población multiparental MAGIC-POP F7

Planta	Contenidos de VTE en frutos					
	α -Tocoferol [ug/gPS]	α -Tocotrienol [ug/gPS]	β -Tocoferol [ug/gPS]	γ -Tocoferol [ug/gPS]	δ -Tocoferol [ug/gPS]	Tocoferol total [ug/gPS]
1	53.25	4.8	2.19	47.85	3.04	106.33
2	2.73	2.77	5.07	NA	NA	7.81
3	6.43	4.28	1.25	11.42	0.7	19.8
4	1.48	8.5	NA	NA	NA	1.48
5	19.32	1.46	1.14	10.37	1.06	31.9
6	58.23	2.89	3.25	17.07	2.92	81.47
7	12.33	14.72	2.77	5.67	42.82	63.59
8	49.81	1.23	3.43	13.95	2.55	69.75
9	39.41	3.13	2	30.83	4.58	76.81
10	12.58	2.15	2.22	22.85	3.83	41.47
11	22.87	0.75	1.36	8.16	1.07	33.46
12	6.01	2.36	0.5	7.55	1.54	15.59
13	26.06	5.68	0.79	44.62	1.77	73.24
14	12.47	1.95	0.93	12.96	1.62	27.98
15	17.5	1.58	0.93	4	0.6	23.03
16	18.19	1.75	1.22	7.12	0.9	27.44
17	11.48	3.49	1.16	12.48	1.45	26.56
18	3.27	2.29	0.28	3.32	0.99	7.87
19	9.57	5.78	1.27	10.19	4.68	25.71
20	7.13	4.02	0.6	7.21	0.78	15.72
21	36.64	1.2	1.65	10.78	1.7	50.77
22	38.56	3.01	1.54	15.42	2.04	57.56
23	8.32	1.09	1.01	7.78	1.45	18.56
24	9.68	2.09	1.69	5.62	1.63	18.62
25	0.77	1.86	0.17	0.39	0.66	2
26	41.99	2.25	1.72	13.81	2.08	59.6
27	9.3	1.22	1.43	6.75	1.24	18.72
28	20.82	3.38	2.09	11.88	1.06	35.85
29	37.49	6.79	2.2	10.77	1	51.45
30	34.85	2.34	3.33	15.25	2.58	56.02
31	7.25	4.91	1.67	6.58	1.76	17.26
32	24.35	2.38	1.43	8.23	1.02	35.03
33	0.78	0.38	0.71	5.36	0.74	7.6
34	31.58	2.59	2.11	6.22	0.2	40.11
35	2.86	2.88	1.51	7.19	0.84	12.4
36	1.19	5.55	0.28	0.7	0.65	2.83
37	2.81	2.61	1.09	3.62	3.53	11.06

38	11.93	2.57	2.56	9.76	2.45	26.69
39	34.55	0.55	2.14	6.68	1.02	44.39
40	12.78	6.19	1.32	17.11	1.76	32.96
41	13.61	2.74	1.33	4.61	1.09	20.65
42	1.48	2.19	0.97	5.95	0.96	9.36
43	11.16	0.75	1.12	7.4	0.92	20.6
44	23.04	0.56	1.27	27.36	1.17	52.84
45	42.49	0.96	1.56	24.24	3.14	71.43
46	42.49	0.96	1.56	24.24	3.14	71.43
47	7.96	1.97	0.63	14.85	2.7	26.14
48	13.46	0.52	0.75	17.13	1.67	33
49	18.33	2.03	0.99	3.33	4.34	26.99
50	26.09	0.83	1.61	5.65	0.98	34.34
51	7.68	0.7	0.49	12.14	1.55	21.85
52	10.99	3.32	0.94	3.44	0.83	16.2
53	35.46	0.65	1.25	28.13	2.39	67.23
54	17.26	3.03	0.88	14.66	0.89	33.68
55	6.39	6.43	0.38	3.27	1.17	11.2
56	3.09	0.38	2.61	0.53	1.63	7.87
57	35.66	2.06	1.22	11.66	1.45	50
58	57.3	4.85	2.84	19.21	2.43	81.78
59	10.05	2.54	0.98	7.07	NA	18.1
60	30.78	2.91	1	15.4	NA	47.18
61	1.23	NA	1.67	0.4	0.6	3.91
62	22.67	1.89	0.86	6.07	0.4	29.99
63	48.03	1.36	2.09	14.35	1.32	65.8
64	4.43	2.12	0.86	3.83	1.1	10.22
65	42.28	1.17	2	8.15	0.69	53.13
66	23.12	1.63	2.65	9.7	NA	35.46
67	27.58	2.68	3.5	21.24	5.19	57.5
68	30.67	0.8	1.74	14.41	2.54	49.37
69	4.02	2.95	0.97	2.32	0.31	7.61
70	30.87	1.96	1.21	8.9	1.28	42.26
71	0.72	0.16	0.79	2.8	1.38	5.69
72	3.97	3.24	0.17	0.93	0.72	5.79
73	18.63	2.85	0.65	7.28	0.51	27.07
74	15.43	1.11	1.13	12.64	1.87	31.07
75	10.91	5.13	0.47	19.99	1.36	32.73
76	25.37	1.01	1.74	10.73	7.61	45.44
77	6.39	1.12	0.51	1.6	0.25	8.74
78	29.35	1.05	1.81	8.79	1.17	41.13
79	20.67	1.68	1.34	21.93	2.79	46.73
80	41.69	1.7	1.54	6.02	0.53	49.78

81	11.2	0.45	0.9	4.57	NA	16.67
82	1.9	1.27	0.15	3.6	0.6	6.25
83	6.36	3.19	1.72	8.83	1.94	18.85
84	39.42	3.15	1.48	17.74	3.01	61.65
85	8.91	1.09	0.86	2.68	0.79	13.24
86	21.78	1.57	1.37	25.64	2.51	51.3
87	4.04	0.5	5.53	2.01	NA	11.59
88	29.41	1.03	1.27	24.15	1.68	56.51
89	38.56	0.79	3.69	2.58	1.57	46.4
90	27.45	2.39	0.83	8.4	NA	36.68
91	13.97	1.02	1.17	5.37	0.96	21.47
92	3.56	1.63	3.3	1.59	1.68	10.13
93	16.82	2.7	0.47	2.22	2.11	21.62
94	3.72	1.34	0.71	3.32	0.58	8.33
95	28.79	3.08	2.42	11.55	3.06	45.82
96	1.41	0.58	0.5	3.63	2.31	7.84
97	37.99	1.14	1.85	15.31	2.36	57.52
98	9.42	2.59	1.16	14.94	2.03	27.55
99	39.32	1.54	2.1	12.13	2.06	55.61
100	27.67	1.3	0.91	29.63	1.46	59.67
101	11.99	2.14	1.75	10.35	2.86	26.96
102	25.66	1.94	1.25	8.88	1.24	37.02
103	17.38	0.73	0.93	21.3	2.48	42.09
104	2.27	0.96	1.04	3.65	1.78	8.75
105	19.64	2.08	1.1	16.01	2.21	38.97
106	4.85	1.38	1.58	9.22	2.31	17.96
107	25.66	0.47	0.66	8.7	0.4	35.42
108	24.98	0.65	0.96	6.89	0.97	33.79
109	21.56	1.32	0.66	0.72	10.03	32.97
110	21.74	0.81	1.45	14.99	4.12	42.31
111	7.96	0.83	0.81	7.04	1.7	17.5
112	35.81	1.72	1.3	13.1	1.41	51.62
113	133.51	4.3	0.71	13.7	1.9	149.83
114	22.08	2.55	1.22	7.19	1.16	31.65
115	16.62	1.01	1.49	8.66	1.8	28.57
116	108.5	2.47	2.11	14.56	2.16	127.32
117	21.91	3.95	1.24	1.36	1.36	25.87
118	18.23	3.52	0.76	11.74	0.1	30.83
119	5.41	0.11	2.62	0.61	2.05	10.69
120	24.54	3.21	1.28	7.66	8.51	42
121	14.33	4.83	0.43	2.37	0.96	18.1
122	60.34	3.72	3.41	11.22	2.66	77.62
123	33.98	6.2	1.42	11.43	1.3	48.13

124	4.52	3.18	0.68	10.2	1.83	17.23
125	33.65	0.68	1.79	22.11	4.02	61.57
126	121.9	0.45	1.99	1.99	11.19	137.08
127	4.76	1.81	1.32	0.94	6.66	13.69
128	34.1	2.66	1.97	11.67	1.64	49.38
129	13.63	0.64	0.93	4.67	1.08	20.31
130	36.68	0.46	2.05	8.33	1.6	48.65
Cer ⁽¹⁾	46.86	0.65	1.58	51.01	2.01	101.46
Cer	47.76	0.56	1.32	42.63	1.64	93.34
Crio ⁽²⁾	34.99	0.11	1.78	24.12	1.69	62.58
Crio	43.62	0.48	1.99	19.95	1.48	67.04
Fer ⁽³⁾	78.07	0.61	3.6	30.72	3.15	115.55
Fer	65.36	NA	3.95	16.91	2.86	89.08
LA014 ⁽⁴⁾	74.13	0.5	2.74	44.42	3.47	124.75
LA014	82.4	0.68	2.74	35.03	2.26	122.43
LA14 ⁽⁵⁾	50.23	0.46	1.69	12.16	0.7	64.78
LA14	80.2	0.63	3.38	63.11	4.72	151.42
LA14	34.74	1.2	0.44	1.98	12.63	49.78
Lev ⁽⁶⁾	58.03	0.11	3.11	18.08	2.08	81.3
Lev	55.48	0.11	3.57	16.82	1.94	77.81
Plov ⁽⁷⁾	4.72	0.03	0.48	2.63	0.74	8.58
Plov	61.2	0.4	1.37	21.86	1.07	85.5
Stup ⁽⁸⁾	84.85	0.11	3.89	34.24	2.59	125.57
Stup	57.26	0.06	2.59	35.75	2.56	98.17
LA014 x Plov ⁽⁹⁾	62.59	1.68	3.86	25.12	3.39	94.96
LA014 x Plov	67.9	0.55	2.92	21.23	2.19	94.25
Fer x LA14 ⁽¹⁰⁾	71.39	0.1	3.38	12.24	1.15	88.16
Fer x LA14	23.55	NA	1.7	8.28	0.86	34.39
Stup x Crio ⁽¹¹⁾	22.87	0.38	1.27	16.48	1.09	41.71
Stup x Crio	76.31	0.79	3.07	32.68	2.8	114.85
Cer x Lev ⁽¹²⁾	71.37	0.33	3.1	37.4	2.84	114.71
Cer x Lev	29.48	0.12	1.61	25.42	2.88	59.4

Datos fenotípicos de los 124 líneas de la MAGIC-POP F7. Contenidos de α - β - γ - δ -tocoferol, α -tocotrienol y tocoferol total medidos por HPLC expresados como [μ g/gPS], μ g de la forma de tococromanol por gramo de tejido seco. Los *parentales* poseen dos replicas biológicas a diferencia de las 124 líneas F7, ⁽¹⁾Cervil, ⁽²⁾Criollo, ⁽³⁾Ferum, ⁽⁴⁾LA0147, ⁽⁵⁾LA1420, ⁽⁶⁾Levovil, ⁽⁷⁾Plovdiv, ⁽⁸⁾Stupicke, ⁽⁹⁾cruza híbrida LA01420 x Plovdiv, ⁽¹⁰⁾cruza híbrida Ferum x LA1420, ⁽¹¹⁾ cruza híbrida Stupicke x Criollo, ⁽¹²⁾ cruza híbrida Cervil x Levovil. NA refiere dato desconocido o faltante.

Tabla S4. Asociaciones identificadas a través de 4 análisis de GWAS con distintos seteos para la determinación de las mismas

Análisis^(a)	QTL^(b)	Condición^(c)	SNP^(d)	pValor^(e)	Intervalo^(f)	Mpb^(g)	N genes^(h)
3	Total	C	S02_46637841	7.20E-07	SL2.50ch02:43085702..47957180	4.9	649
1,2,3	Total	A	S02_49774221	1.70E-06	SL2.50ch02:42354715..52766796	10.4	1383
3	Total	C	S03_05136629	7.60E-05	SL2.50ch03:2147524..56792927	54.6	1517
4	Total	A	S03_05239104	1.90E-05	SL2.50ch03:5136629..5280398	0.1	12
3	Total	E	S04_65402987	5.00E-05	SL2.50ch04:65212262..66025178	0.8	111
2	Total	E	S05_05940022	1.20E-06	SL2.50ch05:3585341...63880865	60.3	1703
1,2,3,4	Total	E	S05_59171718	2.00E-07	SL2.50ch05:3762719..62460829	58.7	1450
3	Total	E	S07_56793907	7.40E-05	SL2.50ch07:4478905..56937221	52.5	832
1,2,3	Total	A	S07_57109419	3.20E-07	SL2.50ch07:56937187..59953976	3.0	231
2	Total	E	S08_00602879	8.90E-05	SL2.50ch08:29999...849696	0.8	1
3	Total	E	S09_04511795	3.90E-05	SL2.50ch09:4415532..4513651	0.1	9
4	Total	A	S09_66591663	9.90E-05	SL2.50ch09:46064192..66735523	20.7	692
2,4	α -T	C	S01_01367023	4.80E-05	SL2.50ch01:1337767...4669442	3.3	329
2	α -T	C	S05_04037126	8.30E-05	SL2.50ch05:4035808...4069911	0.0	4
2	α -T, Total	C	S02_45438825	2.40E-05	SL2.50ch02:43085699..46237131	3.2	418
2,4	α -T, Total	C	S04_66073220	4.20E-06	SL2.50ch04:1125363..66394819	65.3	2488
2,4	α -T3	C	S02_48387492	3.50E-06	SL2.50ch02:41384003...42694399	1.3	178
3	α -T3	A	S05_20443535	9.10E-05	SL2.50ch06:20382279..20576788	0.2	3
2,4	α -T3	C	S06_42560313	2.20E-05	SL2.50ch06:39788857...43770275	4.0	514
2,4	α -T3	A	S09_00835037	4.70E-05	SL2.50ch09:531319..920134	0.4	58
2	α -T3	A	S10_62693839	2.80E-05	SL2.50ch10:62623184..65121083	2.5	363
2,4	α -T3	A	S10_63482478	4.40E-06	SL2.50ch10:63445324..63486547	0.0	4

4	β -T	E	S01_76841190	7.70E-05	SL2.50ch01:76687312..76943377	0.3	21
2	β -T	C	S01_85349943	1.40E-05	SL2.50ch01:85313556...85349971	0.0	6
1,2,3	β -T	C	S01_85997795	2.20E-06	SL2.50ch01:85987757..86155244	0.2	17
1	β -T	C	S01_86198837	1.50E-05	SL2.50ch01:84767788..86330259	1.6	191
4	β -T	E	S01_96492477	2.50E-05	SL2.50ch01:92644554..96492477	3.8	539
4	β -T	C	S02_46038595	3.40E-05	SL2.50ch02:3336890..46970648	43.6	2192
1,2,3	β -T	E	S03_66764638	1.60E-05	SL2.50ch03:66764085..66764866	0.0	1
4	β -T	C	S04_65933475	8.70E-05	SL2.50ch04:66327918..80017664	13.7	24
4	β -T	E	S06_01584491	2.50E-05	SL2.50ch06:1266851..1584811	0.3	40
4	β -T	C	S07_03279240	1.20E-06	SL2.50ch07:3207474..3336890	0.1	5
2	β -T	C	S09_01517398	9.60E-06	SL2.50ch09:1514198...2217770	0.7	74
1	β -T	C	S10_60523403	4.80E-05	SL2.50ch10:59520894..61024199	1.5	193
1	β -T	C	S11_02443617	1.50E-05	SL2.50ch11:2068760..2481305	0.4	40
2	β -T	C	S11_07669410	3.40E-05	SL2.50ch11:4590898...53015083	48.4	1431
2	β -T	C	S11_52844288	1.40E-07	SL2.50ch11:1118868..55733112	54.6	2166
4	β -T	C	S12_09150573	2.30E-08	SL2.50ch12:6682212..47804753	41.1	1168
4	β -T	C	S12_62975968	3.80E-08	SL2.50ch12:62821376..63006395	0.2	21
4	β -T	E	S12_64814575	1.30E-05	SL2.50ch12:64806488..64815085	0.0	2
4	β -T, γ -T	C	S01_02063522	3.50E-05	SL2.50ch01:1809372..2117109	0.3	68
4	γ -T	E	S01_01524332	9.00E-06	SL2.50ch01:348184..2572832	2.2	298
2	γ -T	A	S01_05364660	6.80E-05	SL2.50ch01:5340027..64322523	59.0	815
2,4	γ -T	C	S01_96371008	4.80E-05	SL2.50ch01:67901931...92644554	24.7	1707
4	γ -T	E	S02_45879007	6.80E-05	SL2.50ch02:45841824..45992684	0.2	7
4	γ -T	E	S06_40541416	1.80E-08	SL2.50ch06:40541271..41618939	1.1	130
4	γ -T	A	S07_03845835	5.90E-06	SL2.50ch07:3747303..3976030	0.2	20
4	γ -T	E	S07_64450000	8.10E-06	SL2.50ch07:64287991..67067024	2.8	373
2,4	γ -T	C	S09_67892043	7.80E-05	SL2.50ch09:67687537..67901931	0.2	8
2,4	γ -T	A	S09_72090185	3.30E-05	SL2.50ch09:72090185...72445440	0.4	1

4	γ-T	E	S10_00889234	3.90E-09	SL2.50ch10:794787..910785	0.1	18
4	γ-T	A	S10_51524389	2.50E-05	SL2.50ch10:51336855..54449916	3.1	97
4	γ-T	A	S11_51384690	8.80E-06	SL2.50ch11: 14021263..52192873	38.2	855
4	γ-T	E	S11_54069342	5.80E-06	SL2.50ch11 :53177550..54308337	1.1	125
4	γ-T	A	S11_54288523	2.10E-05	SL2.50ch11:51754613..54290225	2.5	260
2,4	γ-T, Total	A	S05_62738996	1.20E-05	SL2.50ch05:62699279..62748232	0.0	4
4	δ-T	E	S05_00736089	1.40E-05	SL2.50ch05:2874..65349874	65.3	2348
2,4	δ-T	C	S05_63053424	5.40E-05	SL2.50ch05:61892056..63050330	1.2	133
1,2,3,4	δ-T	A	S07_19456301	6.90E-07	SL2.50ch07:4210029..55009164	50.8	741
3	δ-T	A	S09_02160991	9.00E-05	SL2.50ch09:2144983..2194152	0.0	8

^{a)} Numero de análisis en el cual fue determinadas esa asociación al carácter, 1, 2, 3, 4 según su grado de astringencia en el seteo del análisis de GWAS. ^{b)} Caracter/es al cual esta asociado ese polimorfismo identificado. ^{c)} Condición en la cual fue determinada esa asociación, control (C), estrés (E), ambas (A), indicando que no se tubo en cuenta el ambiente para esa asociación. ^{d)} Marcador asociado. ^{e)} Grado de significancia p-valor para ese marcador. ^{f)} Intervalo de confianza para ese marcador determinado en base al LD para esa posición del genoma, ver materiales y métodos sección V.1.6 para mayor detalle. ^{g)}

Tabla S5. Genes identificados a través de los distintos análisis de GWAS y el mapeo de QTLs reportados en la literatura, utilizados para la evaluación de enriquecimiento en la categoría genes relacionados a tococromanos y para la búsqueda de FT sobrerrepresentados

Identificación ^(a)	Ref. Bibliográfica ^(b)	Anotación ^(c)	Locus ^(d)
	Mach 2015	VTE5	Solyc03g071720
4,QTLm	Mach 2015	VTE6	Solyc07g062180
2,4	Hong y col. 2018	COPII	Solyc01g080850
2,4	Hong y col. 2018	CRY1	Solyc04g074180
	Hong y col. 2018	CYP59	Solyc07g066420
QTLm	Hong y col. 2018	DAA1	Solyc03g007760
2,3	Hong y col. 2018	DGAT1-2	Solyc07g040740
	Hong y col. 2018	dnaJ	Solyc09g011880
	Hong y col. 2018	EMB2739	Solyc06g071630
4,QTLm	Hong y col. 2018	GTE8	Solyc07g062660
2,QTLm	Hong y col. 2018	LACS	Solyc11g018580
1,2,4	Hong y col. 2018	LACS	Solyc07g045290
	Hong y col. 2018	LYK2	Solyc09g083210
1,2,3,4	Hong y col. 2018	MGT11	Solyc05g012220
2,3	Hong y col. 2018	NYC1	Solyc07g024000
2,4	Hong y col. 2018	OTP51	Solyc11g066850
	Hong y col. 2018	RbASCH	Solyc09g090050
	Hong y col. 2018	RBP45C	Solyc02g080420
	Hong y col. 2018	Scramblase	Solyc03g006460
	Hong y col. 2018	SINE 2	Solyc03g062700
2,3	Hong y col. 2018	SPP	Solyc01g008820
	Hong y col. 2018	Tic110	Solyc09g031780
	Hong y col. 2018	TPR3	Solyc09g082540
4	Hong y col. 2018	Vib	Solyc12g057120
4	Hong y col. 2018	WRKY46	Solyc02g021680
4	Hong y col. 2018	ZfCCCH19	Solyc02g062360
QTLm	Diepenbrock y col.2017	DAHP	Solyc03g007680
1,2,3,4	Diepenbrock y col.2017	EPSPS(2)	Solyc05g050980
2,4	Diepenbrock y col.2017	ADT/PDH	Solyc11g066890
QTLm	Diepenbrock y col.2017	HGD/HGO	Solyc12g014100
2,4	Diepenbrock y col.2017	HDS(2)	Solyc11g069380
	Diepenbrock y col.2017	LIL3	Solyc08g007180
2,4	Diepenbrock y col.2017	PPH	Solyc01g088090
1,2,3,4	Diepenbrock y col.2017	PDS	Solyc02g080620
	Diepenbrock y col.2017	aroDE	Solyc06g084460
4,QTLm	Diepenbrock y col.2017	SDS	Solyc07g061990
	Guyer y col. 2014	CHL(1)	Solyc06g053980
4	Guyer y col. 2014	CHL(2)	Solyc09g065620
	Guyer y col. 2014	CHL(3)	Solyc09g082600
	Guyer y col. 2014	CHL(4)	Solyc12g005300
	Almeida J y Col, 2011	APT	Solyc06g071550

1,2,3	Almeida J y Col, 2011	CM(1)	Solyc02g088460
2,QTLm	Almeida J y Col, 2011	CM(2)	Solyc11g017240
2,4	Almeida J y Col, 2011	CS(1)	Solyc04g049350
2,4	Almeida J y Col, 2011	CS(2)	Solyc04g009620
2,4	Almeida J y Col, 2011	DAHP	Solyc04g074480
1,2,3,4	Almeida J y Col, 2011	DHQS	Solyc02g083590
	Almeida J y Col, 2011	DQH	Solyc06g084460
	Almeida J y Col, 2011	DXR	Solyc03g114340
2,4	Almeida J y Col, 2011	DXS(1)	Solyc01g067890
2	Almeida J y Col, 2011	DXS(2)	Solyc01g028900
1,2,4	Almeida J y Col, 2011	EPSPS	Solyc01g091190
	Almeida J y Col, 2011	FPGS	Solyc06g051900
	Almeida J y Col, 2011	GGDR	Solyc03g115980
2	Almeida J y Col, 2011	GGPS(1)	Solyc11g011240
2,4	Almeida J y Col, 2011	GGPS(2)	Solyc04g079960
1,2,3	Almeida J y Col, 2011	GGPS(3)	Solyc02g085700
	Almeida J y Col, 2011	GGPS(4)	Solyc09g008920
4	Almeida J y Col, 2011	HDR	Solyc01g109300

^(a) Analisis por el cual fue identificado el gen asociado (Mapeo de QTL), GWAS análisis1,...,4.^(b) Listado de referencia de en los cuales fueron identificados los 94 genes VTEs de esta tabla también reportados en 6 trabajos previos.Tres de ellos en Maiz (Diepenbrock y col. 2017, Hong y col. 2018, Lipka y col. 2013) con lo cual previamente se realizo la busqueda de ortologos mediante OrthoDB v8, quedandonos solo con aquellos que exista alta identidad de secuencia (kriventseva y col.2015). Los otros dos reportes sumados son en tomate (Almeida y Col. 2011, Guyer y col. 2014. ^(c) Identificador del Locus. ^(d) Anotacion de acuerdo a ITAG2.4 sobre SL2.50 genoma.

Tabla S6. Ensayo de apilamiento de dos loci. F2 cruza (IL9-2-6-1 X IL3-1)

			A) Contenidos VTE frutos					
Planta ¹	Ambiente ²	Genotipo ³	α-Tocoferol	β-Tocoferol	γ-Tocoferol	δ-Tocoferol	Total	α+γ/β+δ
P-10	Castelar	IL 3-1	7.39	0.22	1.46	0.26	9.33	18.47
P-5	Castelar	IL 3-1	7.98	0.31	2.84	0.55	11.67	12.59
P-3	Castelar	IL 3-1	4.34	0.11	1.42	0.24	6.12	16.29
P-10	Mendoza	IL 3-1	10.72	0.43	0.39	0.07	11.61	22.31
P-5	Mendoza	IL 3-1	10.39	0.55	0.24	0.07	11.25	17.16
P-3	Mendoza	IL 3-1	8.49	0.41	0.61	0.10	9.62	17.73
P-22	Castelar	IL 9-2-6-1	7.33	0.16	0.55	0.06	8.10	36.67
P-23	Castelar	IL 9-2-6-1	5.68	0.16	0.13	0.01	5.99	32.48
P-27	Castelar	IL 9-2-6-1	6.08	0.10	0.92	0.18	7.28	25.00
P-22	Mendoza	IL 9-2-6-1	7.75	0.36	1.60	0.17	9.89	17.59
P-23	Mendoza	IL 9-2-6-1	6.73	0.38	0.49	0.08	7.68	15.73
P-27	Mendoza	IL 9-2-6-1	7.98	0.38	1.16	0.14	9.66	17.54
P-1	Castelar	M82	8.80	0.25	2.45	0.50	12.00	15.02
P-5	Castelar	M82	6.07	0.10	0.45	0.05	6.67	43.50
P-7	Castelar	M82	5.88	0.25	2.04	0.49	8.65	10.81
P-1	Mendoza	M82	8.67	0.39	0.85	0.13	10.04	18.32
P-5	Mendoza	M82	9.26	0.40	0.16	0.04	9.86	21.08
P-7	Mendoza	M82	11.33	0.51	0.83	0.14	12.80	18.81
R-47	Castelar	R1	8.68	0.18	1.72	0.22	10.80	25.90
R-56	Castelar	R1	7.86	0.17	1.16	0.15	9.34	28.26
R-75	Castelar	R1	7.97	0.18	2.52	0.02	10.69	53.96
R-47	Mendoza	R1	9.51	0.40	0.34	0.07	10.32	20.92
R-56	Mendoza	R1	7.23	0.34	0.89	0.13	8.60	17.18
R-75	Mendoza	R1	12.02	0.41	1.33	0.13	13.90	24.51
R-78	Castelar	R2	3.19	0.12	0.67	0.07	4.05	21.05
R-86	Castelar	R2	4.57	0.13	0.93	0.10	5.73	23.61
R-87	Castelar	R2	7.86	0.13	1.27	0.10	9.35	40.33
R-78	Mendoza	R2	10.07	0.44	0.87	0.10	11.48	20.44
R-86	Mendoza	R2	9.72	0.47	0.33	0.07	10.59	18.56
R-87	Mendoza	R2	9.42	0.34	0.76	0.08	10.60	24.18
R-49	Castelar	R3	6.02	0.14	0.51	0.07	6.74	31.62
R-110	Castelar	R3	8.15	0.20	0.89	0.12	9.36	28.21
R-119	Castelar	R3	7.50	0.22	1.50	0.20	9.41	21.66
R-49	Mendoza	R3	6.50	0.45	0.20	0.07	7.22	13.02
R-110	Mendoza	R3	10.99	0.49	0.31	0.06	11.85	20.42
R-119	Mendoza	R3	12.62	0.53	0.33	0.08	13.55	21.36
R-42	Castelar	R4	10.00	0.16	1.21	0.13	11.50	39.31
R-90	Castelar	R4	8.93	0.17	1.94	0.26	11.29	25.49
R-118	Castelar	R4	6.10	0.14	0.97	0.11	7.33	27.42
R-42	Mendoza	R4	11.31	0.59	0.16	0.04	12.11	18.04
R-90	Mendoza	R4	9.48	0.47	0.26	0.07	10.28	17.93
R-118	Mendoza	R4	12.47	0.41	1.21	0.14	14.22	25.09
R-83	Castelar	R5	6.52	0.12	0.13	0.01	6.77	52.24

R-88	Castelar	R5	8.62	0.21	0.78	0.11	9.72	29.68
R-100	Castelar	R5	4.31	0.10	0.51	0.04	4.95	34.60
R-83	Mendoza	R5	8.35	0.33	0.16	0.05	8.89	22.39
R-88	Mendoza	R5	9.54	0.44	0.29	0.08	10.35	18.93
R-100	Mendoza	R5	9.08	0.32	1.06	0.11	10.57	23.62
R-101	Castelar	R6	5.32	0.24	0.34	0.06	5.96	18.44
R-111	Castelar	R6	7.82	0.14	1.76	0.20	9.92	27.73
R-113	Castelar	R6	3.85	0.21	0.78	0.08	4.92	15.59
R-101	Mendoza	R6	11.21	0.48	0.26	0.09	12.04	20.43
R-111	Mendoza	R6	8.29	0.44	0.17	0.06	8.96	17.04
R-113	Mendoza	R6	9.86	0.47	1.04	0.13	11.50	18.15
R-41	Castelar	R7	8.62	0.29	0.60	0.10	9.61	23.52
R-57	Castelar	R7	9.37	0.24	1.11	0.16	10.87	26.54
R-93	Castelar	R7	6.60	0.13	0.19	0.01	6.93	48.81
R-41	Mendoza	R7	8.58	0.50	0.58	0.15	9.81	14.24
R-57	Mendoza	R7	12.04	0.46	1.13	0.18	13.81	20.67
R-93	Mendoza	R7	8.49	0.41	0.61	0.10	9.62	17.73
R-70	Castelar	R8	9.66	0.21	0.73	0.09	10.69	34.61
R-89	Castelar	R8	6.61	0.29	0.24	0.03	7.18	20.91
R-94	Castelar	R8	9.27	0.18	1.89	0.29	11.64	23.64
R-70	Mendoza	R8	8.35	0.42	0.85	0.14	9.76	16.24
R-89	Mendoza	R8	8.79	0.47	0.22	0.06	9.54	16.88
R-94	Mendoza	R8	14.78	0.55	0.79	0.16	16.29	21.77
R-96	Castelar	R9	10.38	0.20	1.41	0.19	12.18	30.31
R-102	Castelar	R9	8.64	0.31	0.42	0.08	9.46	22.99
R-96(2)	Castelar	R9	11.90	0.21	2.01	0.24	14.36	30.47
R-96	Mendoza	R9	10.49	0.45	0.53	0.11	11.58	19.66
R-102	Mendoza	R9	11.50	0.59	0.59	0.16	12.84	16.10
R-96(2)	Mendoza	R9	11.75	0.53	0.11	0.04	12.42	21.08
			B) Contenidos VTE hojas					
Planta ¹	Ambiente ²	Genotipo ³	α -Tocoferol	β -Tocoferol	γ -Tocoferol	δ -Tocoferol	Total	$\alpha+\gamma/\beta+\delta$
P-10	Castelar	IL 3-1	67.43	0.27	2.55	0.37	70.61	109.34
P-5	Castelar	IL 3-1	36.62	0.15	0.49	0.25	37.51	92.78
P-3	Castelar	IL 3-1	32.14	0.28	0.77	0.40	33.58	48.40
P-10	Mendoza	IL 3-1	75.44	3.00	1.20	0.66	80.30	20.95
P-5	Mendoza	IL 3-1	92.32	2.89	1.14	0.39	96.74	28.48
P-3	Mendoza	IL 3-1	109.93	2.88	0.87	0.47	114.15	33.11
P-22	Castelar	IL 9-2-6-1	35.58	0.50	1.01	0.61	37.71	32.96
P-23	Castelar	IL 9-2-6-1	40.46	0.36	0.93	0.19	41.93	75.25
P-27	Castelar	IL 9-2-6-1	58.00	0.20	0.57	0.15	58.93	167.34
P-22	Mendoza	IL 9-2-6-1	112.77	3.04	1.07	0.43	117.31	32.76
P-23	Mendoza	IL 9-2-6-1	118.26	3.31	1.33	0.83	123.74	28.84
P-27	Mendoza	IL 9-2-6-1	82.49	3.04	1.09	3.02	89.63	13.81
P-1	Castelar	M82	56.21	0.64	1.17	0.18	58.20	69.98
P-5	Castelar	M82	32.97	0.66	1.01	0.57	35.21	27.63
P-7	Castelar	M82	23.90	0.24	0.85	0.18	25.17	58.93
P-1	Mendoza	M82	61.32	2.68	0.88	0.50	65.37	19.57

P-5	Mendoza	M82	74.88	3.16	1.23	0.41	79.68	21.34
P-7	Mendoza	M82	75.42	2.75	1.10	0.44	79.71	24.02
R-47	Castelar	R1	30.74	0.24	0.77	0.44	32.19	46.34
R-56	Castelar	R1	64.48	2.75	0.96	0.66	68.85	19.17
R-75	Castelar	R1	66.56	0.66	0.65	0.20	68.07	78.15
R-47	Mendoza	R1	89.34	2.66	0.92	0.47	93.40	28.80
R-56	Mendoza	R1	31.97	0.27	0.43	0.30	32.98	56.84
R-75	Mendoza	R1	97.72	3.12	0.98	0.45	102.27	27.67
R-78	Castelar	R2	28.80	0.50	0.31	0.00	29.60	58.22
R-86	Castelar	R2	19.90	0.27	0.43	0.20	20.80	43.26
R-87	Castelar	R2	26.97	0.54	1.96	0.09	29.55	45.92
R-78	Mendoza	R2	105.80	3.30	1.06	0.59	110.75	27.46
R-86	Mendoza	R2	115.54	6.09	2.32	0.91	124.87	16.84
R-87	Mendoza	R2	67.80	2.34	0.81	0.39	71.35	25.16
R-49	Castelar	R3	40.82	0.34	0.80	0.41	42.35	55.49
R-110	Castelar	R3	58.85	0.69	0.67	0.59	60.79	46.50
R-119	Castelar	R3	31.89	3.32	2.96	0.00	38.17	10.50
R-49	Mendoza	R3	111.87	3.05	1.01	0.72	116.65	29.96
R-110	Mendoza	R3	97.00	2.82	1.01	0.44	101.27	30.07
R-119	Mendoza	R3	110.79	4.10	1.22	0.70	116.81	23.34
R-42	Castelar	R4	34.57	1.49	0.92	0.52	37.49	17.66
R-90	Castelar	R4	32.16	0.43	0.62	0.42	33.63	38.56
R-118	Castelar	R4	114.35	1.07	2.34	0.48	118.23	75.28
R-42	Mendoza	R4	130.44	3.61	1.26	0.76	136.07	30.15
R-90	Mendoza	R4	100.33	2.70	0.98	0.55	104.56	31.12
R-118	Mendoza	R4	138.55	2.73	0.94	1.10	143.32	36.46
R-83	Castelar	R5	58.37	0.62	1.13	0.80	60.93	41.90
R-88	Castelar	R5	21.50	0.24	1.40	0.09	23.23	69.39
R-100	Castelar	R5	34.02	0.51	0.65	0.35	35.53	40.31
R-83	Mendoza	R5	109.10	3.46	0.96	0.71	114.22	26.39
R-88	Mendoza	R5	100.61	3.33	1.49	0.88	106.32	24.24
R-100	Mendoza	R5	81.67	1.93	0.73	0.32	84.65	36.54
R-101	Castelar	R6	20.94	0.29	1.60	0.04	22.87	68.30
R-111	Castelar	R6	21.71	0.18	1.06	0.00	22.95	126.50
R-113	Castelar	R6	33.19	0.51	0.95	0.30	34.95	42.15
R-101	Mendoza	R6	106.56	2.52	0.83	0.45	110.36	36.16
R-111	Mendoza	R6	127.90	3.28	1.03	0.65	132.87	32.75
R-113	Mendoza	R6	59.42	3.82	1.15	0.50	64.89	14.03
R-41	Castelar	R7	44.54	0.31	1.39	0.21	46.45	88.33
R-57	Castelar	R7	32.87	0.71	0.74	0.00	34.31	47.34
R-93	Castelar	R7	46.35	0.44	1.04	0.15	47.99	80.32
R-41	Mendoza	R7	93.31	4.81	1.57	0.63	100.33	17.44
R-57	Mendoza	R7	101.76	2.53	0.89	0.48	105.66	34.10
R-93	Mendoza	R7	89.79	3.86	1.72	0.50	95.88	20.96
R-70	Castelar	R8	62.10	1.35	1.15	0.55	65.15	33.29
R-89	Castelar	R8	63.39	0.41	0.56	0.10	64.47	125.39
R-94	Castelar	R8	32.94	2.07	1.92	0.16	37.09	15.63
R-70	Mendoza	R8	167.06	4.64	1.57	0.91	174.18	30.36

R-89	Mendoza	R8	74.04	2.41	0.89	0.58	77.91	25.08
R-94	Mendoza	R8	59.49	2.08	0.87	0.41	62.86	24.20
R-96	Castelar	R9	27.10	0.36	0.75	0.20	28.41	49.73
R-102	Castelar	R9	49.10	0.45	1.03	0.84	51.42	38.86
R-96(2)	Castelar	R9	34.73	0.17	0.64	0.29	35.82	76.89
R-96	Mendoza	R9	123.71	4.46	1.54	0.62	130.33	24.66
R-102	Mendoza	R9	148.76	2.69	1.23	0.00	152.68	55.76
R-96(2)	Mendoza	R9	130.58	4.12	1.72	0.59	137.01	28.07
			C) Log ₂ Expresión relativa frutos					
Planta ¹	Ambiente ²	Genotipo ³	VTE3(1)L ⁴	VTE3(1)P ⁵	VTE3(2)L ⁶	VTE3(2)P ⁷	ΣExp Alelos ⁸	
P-10	Castelar	IL 3-1	-4.47	NA	NA	0.19	0.24	
P-5	Castelar	IL 3-1	-3.73	NA	NA	0.98	1.03	
P-3	Castelar	IL 3-1	-4.25	NA	NA	0.91	0.94	
P-10	Mendoza	IL 3-1	-2.28	NA	NA	2.21	2.28	
P-5	Mendoza	IL 3-1	-2.14	NA	NA	2.40	2.46	
P-3	Mendoza	IL 3-1	-2.63	NA	NA	2.34	2.39	
P-22	Castelar	IL 9-2-6-1	NA	-4.43	0.13	NA	0.19	
P-23	Castelar	IL 9-2-6-1	NA	-2.89	1.32	NA	1.40	
P-27	Castelar	IL 9-2-6-1	NA	-3.88	-0.57	NA	-0.43	
P-22	Mendoza	IL 9-2-6-1	NA	-3.14	0.59	NA	0.70	
P-23	Mendoza	IL 9-2-6-1	NA	-1.30	2.12	NA	2.25	
P-27	Mendoza	IL 9-2-6-1	NA	-2.34	0.19	NA	0.42	
P-1	Castelar	M82	-1.79	NA	-0.29	NA	0.15	
P-5	Castelar	M82	-2.12	NA	0.36	NA	0.60	
P-7	Castelar	M82	-2.25	NA	0.66	NA	0.84	
P-1	Mendoza	M82	-0.30	NA	1.15	NA	1.60	
P-5	Mendoza	M82	-1.79	NA	0.59	NA	0.84	
P-7	Mendoza	M82	-1.80	NA	0.98	NA	1.17	
R-47	Castelar	R1	NA	-0.74	2.33	NA	2.49	
R-56	Castelar	R1	NA	-1.65	0.50	NA	0.80	
R-75	Castelar	R1	NA	-2.19	1.17	NA	1.30	
R-47	Mendoza	R1	NA	-0.18	1.67	NA	2.03	
R-56	Mendoza	R1	NA	-0.66	1.25	NA	1.59	
R-75	Mendoza	R1	NA	-1.44	1.15	NA	1.37	
R-78	Castelar	R2	NA	-2.92	0.68	0.97	1.88	
R-86	Castelar	R2	NA	-3.14	0.12	0.34	1.30	
R-87	Castelar	R2	NA	-2.57	0.58	0.35	1.56	
R-78	Mendoza	R2	NA	-1.10	2.17	2.50	3.41	
R-86	Mendoza	R2	NA	-1.41	1.55	1.81	2.77	
R-87	Mendoza	R2	NA	-1.14	1.71	1.55	2.73	
R-49	Castelar	R3	-3.13	-1.50	2.95	NA	3.03	
R-110	Castelar	R3	-2.56	-2.61	2.89	NA	2.95	
R-119	Castelar	R3	-1.22	-2.70	2.43	NA	2.58	
R-49	Mendoza	R3	-1.09	-1.09	2.33	NA	2.58	
R-110	Mendoza	R3	-1.08	-0.86	3.57	NA	3.69	
R-119	Mendoza	R3	-2.33	-1.65	2.41	NA	2.55	
R-42	Castelar	R4	-2.83	-3.45	-0.25	0.25	1.18	

R-90	Castelar	R4	-3.01	-3.62	-0.39	-0.40	0.79
R-118	Castelar	R4	-2.44	-3.47	-0.70	-0.19	0.82
R-42	Mendoza	R4	-1.24	-1.61	0.56	1.76	2.49
R-90	Mendoza	R4	-1.50	-1.86	0.35	1.05	1.99
R-118	Mendoza	R4	-1.22	-1.98	-0.22	0.99	1.82
R-83	Castelar	R5	NA	-0.92	NA	1.69	1.91
R-88	Castelar	R5	NA	-0.80	NA	1.57	1.82
R-100	Castelar	R5	NA	0.00	NA	1.85	2.20
R-83	Mendoza	R5	NA	1.59	NA	3.93	4.19
R-88	Mendoza	R5	NA	1.66	NA	3.78	4.08
R-100	Mendoza	R5	NA	2.17	NA	3.81	4.21
R-101	Castelar	R6	-1.85	-2.37	NA	-0.34	0.33
R-111	Castelar	R6	-2.73	-2.72	NA	-0.18	0.24
R-113	Castelar	R6	-3.39	-2.61	NA	-0.24	0.15
R-101	Mendoza	R6	0.63	0.11	NA	1.86	2.64
R-111	Mendoza	R6	-0.85	-0.25	NA	1.98	2.42
R-113	Mendoza	R6	-0.46	-0.37	NA	1.69	2.24
R-41	Castelar	R7	-3.75	NA	-1.11	NA	-0.89
R-57	Castelar	R7	-3.98	NA	-1.23	NA	-1.03
R-93	Castelar	R7	-5.03	NA	-1.47	NA	-1.36
R-41	Mendoza	R7	-1.85	NA	0.67	NA	0.90
R-57	Mendoza	R7	-1.99	NA	0.64	NA	0.86
R-93	Mendoza	R7	-2.75	NA	0.65	NA	0.78
R-70	Castelar	R8	-1.19	NA	1.45	0.91	2.34
R-89	Castelar	R8	-0.34	NA	0.97	0.98	2.24
R-94	Castelar	R8	-2.33	NA	0.82	0.86	1.92
R-70	Mendoza	R8	-0.90	NA	1.67	1.22	2.60
R-89	Mendoza	R8	-0.62	NA	0.69	0.75	1.98
R-94	Mendoza	R8	-2.07	NA	0.99	1.09	2.12
R-96	Castelar	R9	-3.49	NA	NA	0.46	0.55
R-102	Castelar	R9	-2.70	NA	NA	0.76	0.88
R-96(2)	Castelar	R9	-2.98	NA	NA	0.54	0.66
R-96	Mendoza	R9	-2.17	NA	NA	1.66	1.76
R-102	Mendoza	R9	-2.18	NA	NA	1.18	1.31
R-96(2)	Mendoza	R9	-2.01	NA	NA	1.40	1.53
			D) Log₂ Expresión relativa Hojas				
Planta ¹	Ambiente ²	Genotipo ³	VTE3(1)L ⁴	VTE3(1)P ⁵	VTE3(2)L ⁶	VTE3(2)P ⁷	ΣExp Alelos ⁸
P-10	Castelar	IL 3-1	-4.47	NA	NA	0.19	0.24
P-5	Castelar	IL 3-1	-3.73	NA	NA	0.98	1.03
P-3	Castelar	IL 3-1	-4.25	NA	NA	0.91	0.94
P-10	Mendoza	IL 3-1	-2.28	NA	NA	2.21	2.28
P-5	Mendoza	IL 3-1	-2.14	NA	NA	2.40	2.46
P-3	Mendoza	IL 3-1	-2.63	NA	NA	2.34	2.39
P-22	Castelar	IL 9-2-6-1	NA	-4.43	0.13	NA	0.19
P-23	Castelar	IL 9-2-6-1	NA	-2.89	1.32	NA	1.40
P-27	Castelar	IL 9-2-6-1	NA	-3.88	-0.57	NA	-0.43
P-22	Mendoza	IL 9-2-6-1	NA	-3.14	0.59	NA	0.70

P-23	Mendoza	IL 9-2-6-1	NA	-1.30	2.12	NA	2.25
P-27	Mendoza	IL 9-2-6-1	NA	-2.34	0.19	NA	0.42
P-1	Castelar	M82	-1.79	NA	-0.29	NA	0.15
P-5	Castelar	M82	-2.12	NA	0.36	NA	0.60
P-7	Castelar	M82	-2.25	NA	0.66	NA	0.84
P-1	Mendoza	M82	-0.30	NA	1.15	NA	1.60
P-5	Mendoza	M82	-1.79	NA	0.59	NA	0.84
P-7	Mendoza	M82	-1.80	NA	0.98	NA	1.17
R-47	Castelar	R1	NA	-0.74	2.33	NA	2.49
R-56	Castelar	R1	NA	-1.65	0.50	NA	0.80
R-75	Castelar	R1	NA	-2.19	1.17	NA	1.30
R-47	Mendoza	R1	NA	-0.18	1.67	NA	2.03
R-56	Mendoza	R1	NA	-0.66	1.25	NA	1.59
R-75	Mendoza	R1	NA	-1.44	1.15	NA	1.37
R-78	Castelar	R2	NA	-2.92	0.68	0.97	1.88
R-86	Castelar	R2	NA	-3.14	0.12	0.34	1.30
R-87	Castelar	R2	NA	-2.57	0.58	0.35	1.56
R-78	Mendoza	R2	NA	-1.10	2.17	2.50	3.41
R-86	Mendoza	R2	NA	-1.41	1.55	1.81	2.77
R-87	Mendoza	R2	NA	-1.14	1.71	1.55	2.73
R-49	Castelar	R3	-3.13	-1.50	2.95	NA	3.03
R-110	Castelar	R3	-2.56	-2.61	2.89	NA	2.95
R-119	Castelar	R3	-1.22	-2.70	2.43	NA	2.58
R-49	Mendoza	R3	-1.09	-1.09	2.33	NA	2.58
R-110	Mendoza	R3	-1.08	-0.86	3.57	NA	3.69
R-119	Mendoza	R3	-2.33	-1.65	2.41	NA	2.55
R-42	Castelar	R4	-2.83	-3.45	-0.25	0.25	1.18
R-90	Castelar	R4	-3.01	-3.62	-0.39	-0.40	0.79
R-118	Castelar	R4	-2.44	-3.47	-0.70	-0.19	0.82
R-42	Mendoza	R4	-1.24	-1.61	0.56	1.76	2.49
R-90	Mendoza	R4	-1.50	-1.86	0.35	1.05	1.99
R-118	Mendoza	R4	-1.22	-1.98	-0.22	0.99	1.82
R-83	Castelar	R5	NA	-0.92	NA	1.69	1.91
R-88	Castelar	R5	NA	-0.80	NA	1.57	1.82
R-100	Castelar	R5	NA	0.00	NA	1.85	2.20
R-83	Mendoza	R5	NA	1.59	NA	3.93	4.19
R-88	Mendoza	R5	NA	1.66	NA	3.78	4.08
R-100	Mendoza	R5	NA	2.17	NA	3.81	4.21
R-101	Castelar	R6	-1.85	-2.37	NA	-0.34	0.33
R-111	Castelar	R6	-2.73	-2.72	NA	-0.18	0.24
R-113	Castelar	R6	-3.39	-2.61	NA	-0.24	0.15
R-101	Mendoza	R6	0.63	0.11	NA	1.86	2.64
R-111	Mendoza	R6	-0.85	-0.25	NA	1.98	2.42
R-113	Mendoza	R6	-0.46	-0.37	NA	1.69	2.24
R-41	Castelar	R7	-3.75	NA	-1.11	NA	-0.89
R-57	Castelar	R7	-3.98	NA	-1.23	NA	-1.03
R-93	Castelar	R7	-5.03	NA	-1.47	NA	-1.36
R-41	Mendoza	R7	-1.85	NA	0.67	NA	0.90

R-57	Mendoza	R7	-1.99	NA	0.64	NA	0.86
R-93	Mendoza	R7	-2.75	NA	0.65	NA	0.78
R-70	Castelar	R8	-1.19	NA	1.45	0.91	2.34
R-89	Castelar	R8	-0.34	NA	0.97	0.98	2.24
R-94	Castelar	R8	-2.33	NA	0.82	0.86	1.92
R-70	Mendoza	R8	-0.90	NA	1.67	1.22	2.60
R-89	Mendoza	R8	-0.62	NA	0.69	0.75	1.98
R-94	Mendoza	R8	-2.07	NA	0.99	1.09	2.12
R-96	Castelar	R9	-3.49	NA	NA	0.46	0.55
R-102	Castelar	R9	-2.70	NA	NA	0.76	0.88
R-96(2)	Castelar	R9	-2.98	NA	NA	0.54	0.66
R-96	Mendoza	R9	-2.17	NA	NA	1.66	1.76
R-102	Mendoza	R9	-2.18	NA	NA	1.18	1.31
R-96(2)	Mendoza	R9	-2.01	NA	NA	1.40	1.53
E) Nivel metilación región promotora (McrBc) frutos							
Planta ¹	Ambiente ²	Genotipo ³	VTE3(1)L ⁴	VTE3(1)P ⁵	VTE3(2)L ⁶	VTE3(2)P ⁷	MM ⁹
P-10	Castelar	IL 3-1	52.95	NA	NA	97.07	75.01
P-5	Castelar	IL 3-1	0.00	NA	NA	96.89	48.44
P-3	Castelar	IL 3-1	60.06	NA	NA	99.12	79.59
P-10	Mendoza	IL 3-1	12.82	NA	NA	100.00	56.41
P-5	Mendoza	IL 3-1	7.35	NA	NA	98.56	52.95
P-3	Mendoza	IL 3-1	18.28	NA	NA	98.81	58.55
P-22	Castelar	IL 9-2-6-1	NA	31.99	99.03	NA	65.51
P-23	Castelar	IL 9-2-6-1	NA	0.00	98.90	NA	49.45
P-27	Castelar	IL 9-2-6-1	NA	63.98	99.16	NA	81.57
P-22	Mendoza	IL 9-2-6-1	NA	0.00	97.24	NA	48.62
P-23	Mendoza	IL 9-2-6-1	NA	69.17	98.79	NA	83.98
P-27	Mendoza	IL 9-2-6-1	NA	0.35	99.31	NA	49.83
P-1	Castelar	M82	97.60	NA	97.51	NA	97.55
P-5	Castelar	M82	97.16	NA	98.47	NA	97.82
P-7	Castelar	M82	60.58	NA	98.00	NA	79.29
P-1	Mendoza	M82	0.00	NA	98.16	NA	49.08
P-5	Mendoza	M82	24.06	NA	99.89	NA	61.97
P-7	Mendoza	M82	12.03	NA	99.02	NA	55.53
R-47	Castelar	R1	NA	73.93	96.53	NA	85.23
R-56	Castelar	R1	NA	26.76	97.13	NA	61.94
R-75	Castelar	R1	NA	81.52	96.60	NA	89.06
R-47	Mendoza	R1	NA	7.38	97.17	NA	52.27
R-56	Mendoza	R1	NA	58.00	95.94	NA	76.97
R-75	Mendoza	R1	NA	32.69	96.55	NA	64.62
R-78	Castelar	R2	NA	70.74	96.64	91.45	86.28
R-86	Castelar	R2	NA	91.57	97.42	90.91	93.30
R-87	Castelar	R2	NA	25.81	97.42	92.48	71.90
R-78	Mendoza	R2	NA	59.78	97.41	90.07	82.42
R-86	Mendoza	R2	NA	0.00	97.69	86.03	61.24
R-87	Mendoza	R2	NA	56.20	97.43	85.59	79.74
R-49	Castelar	R3	73.47	28.73	28.49	NA	43.56

R-110	Castelar	R3	87.01	70.14	96.33	NA	84.49
R-119	Castelar	R3	45.21	89.73	37.80	NA	57.58
R-49	Mendoza	R3	80.62	74.94	75.37	NA	76.98
R-110	Mendoza	R3	80.01	59.17	19.14	NA	52.77
R-119	Mendoza	R3	83.06	69.36	18.30	NA	56.91
R-42	Castelar	R4	16.27	36.47	96.32	96.29	61.34
R-90	Castelar	R4	99.65	83.34	98.13	96.23	94.34
R-118	Castelar	R4	99.46	1.80	66.09	94.75	65.53
R-42	Mendoza	R4	93.86	0.00	67.60	97.70	64.79
R-90	Mendoza	R4	1.90	40.32	97.74	98.22	59.55
R-118	Mendoza	R4	56.05	17.66	37.05	88.31	49.76
R-83	Castelar	R5	NA	18.99	NA	92.74	55.86
R-88	Castelar	R5	NA	25.40	NA	89.40	57.40
R-100	Castelar	R5	NA	0.00	NA	91.72	45.86
R-83	Mendoza	R5	NA	0.00	NA	91.19	45.60
R-88	Mendoza	R5	NA	83.21	NA	92.07	87.64
R-100	Mendoza	R5	NA	0.00	NA	90.47	45.24
R-101	Castelar	R6	79.43	70.48	NA	0.00	49.97
R-111	Castelar	R6	49.39	36.81	NA	0.00	28.73
R-113	Castelar	R6	33.63	80.75	NA	0.00	38.13
R-101	Mendoza	R6	96.57	43.64	NA	0.00	46.74
R-111	Mendoza	R6	94.84	92.29	NA	0.00	62.37
R-113	Mendoza	R6	90.55	67.96	NA	0.00	52.84
R-41	Castelar	R7	96.23	NA	98.73	NA	97.48
R-57	Castelar	R7	86.70	NA	98.22	NA	92.46
R-93	Castelar	R7	0.00	NA	98.51	NA	49.26
R-41	Mendoza	R7	85.50	NA	95.87	NA	90.68
R-57	Mendoza	R7	77.53	NA	93.92	NA	85.73
R-93	Mendoza	R7	0.00	NA	97.10	NA	48.55
R-70	Castelar	R8	31.31	NA	98.41	90.41	73.38
R-89	Castelar	R8	56.18	NA	98.03	89.71	81.31
R-94	Castelar	R8	43.75	NA	98.22	100.00	80.66
R-70	Mendoza	R8	16.20	NA	98.50	95.01	69.90
R-89	Mendoza	R8	10.99	NA	98.29	95.65	68.31
R-94	Mendoza	R8	41.75	NA	98.34	94.35	78.15
R-96	Castelar	R9	62.31	NA	NA	94.76	78.53
R-102	Castelar	R9	13.26	NA	NA	97.37	55.31
R-96(2)	Castelar	R9	4.25	NA	NA	97.68	50.96
R-96	Mendoza	R9	0.00	NA	NA	96.85	48.43
R-102	Mendoza	R9	32.06	NA	NA	97.10	64.58
R-96(2)	Mendoza	R9	21.90	NA	NA	95.78	58.84
F) Nivel metilación región promotora (McrBc) hojas							
Planta ¹	Ambiente ²	Genotipo ³	VTE3(1)L ⁴	VTE3(1)P ⁵	VTE3(2)L ⁶	VTE3(2)P ⁷	MM ⁹
P-10	Castelar	IL 3-1	0.00	NA	NA	6.58	3.29
P-5	Castelar	IL 3-1	0.00	NA	NA	0.00	0.00
P-3	Castelar	IL 3-1	0.00	NA	NA	0.00	0.00
P-10	Mendoza	IL 3-1	9.65	NA	NA	0.84	5.25

P-5	Mendoza	IL 3-1	4.60	NA	NA	9.43	7.01
P-3	Mendoza	IL 3-1	0.00	NA	NA	10.14	5.07
P-22	Castelar	IL 9-2-6-1	NA	14.32	11.73	NA	13.02
P-23	Castelar	IL 9-2-6-1	NA	0.00	0.00	NA	0.00
P-27	Castelar	IL 9-2-6-1	NA	4.74	0.00	NA	2.37
P-22	Mendoza	IL 9-2-6-1	NA	0.00	11.62	NA	5.81
P-23	Mendoza	IL 9-2-6-1	NA	0.00	1.99	NA	1.00
P-27	Mendoza	IL 9-2-6-1	NA	0.00	4.82	NA	2.41
P-1	Castelar	M82	1.92	NA	0.00	NA	0.96
P-5	Castelar	M82	4.85	NA	0.00	NA	2.43
P-7	Castelar	M82	1.26	NA	3.00	NA	2.13
P-1	Mendoza	M82	5.20	NA	0.00	NA	2.60
P-5	Mendoza	M82	8.35	NA	0.00	NA	4.17
P-7	Mendoza	M82	1.54	NA	0.00	NA	0.77
R-47	Castelar	R1	NA	0.00	0.00	NA	0.00
R-56	Castelar	R1	NA	1.98	15.68	NA	8.83
R-75	Castelar	R1	NA	8.74	0.66	NA	4.70
R-47	Mendoza	R1	NA	0.17	16.25	NA	8.21
R-56	Mendoza	R1	NA	0.00	5.34	NA	2.67
R-75	Mendoza	R1	NA	11.53	0.00	NA	5.77
R-78	Castelar	R2	NA	12.61	8.98	0.00	7.20
R-86	Castelar	R2	NA	0.00	0.00	0.00	0.00
R-87	Castelar	R2	NA	0.00	7.25	4.13	3.79
R-78	Mendoza	R2	NA	12.37	9.10	0.00	7.15
R-86	Mendoza	R2	NA	0.33	0.00	0.00	0.11
R-87	Mendoza	R2	NA	13.53	12.31	0.00	8.61
R-49	Castelar	R3	0.00	0.00	5.53	NA	1.84
R-110	Castelar	R3	0.00	0.00	17.46	NA	5.82
R-119	Castelar	R3	0.00	0.00	4.28	NA	1.43
R-49	Mendoza	R3	5.21	10.95	0.46	NA	5.54
R-110	Mendoza	R3	0.00	0.00	4.95	NA	1.65
R-119	Mendoza	R3	0.00	0.00	18.15	NA	6.05
R-42	Castelar	R4	8.79	100.00	20.00	0.00	32.20
R-90	Castelar	R4	8.43	100.00	0.00	8.33	29.19
R-118	Castelar	R4	0.00	78.10	18.92	0.00	24.25
R-42	Mendoza	R4	18.79	0.00	0.00	4.81	5.90
R-90	Mendoza	R4	0.28	0.00	0.00	0.00	0.07
R-118	Mendoza	R4	0.00	0.00	0.77	0.00	0.19
R-83	Castelar	R5	NA	0.81	NA	0.00	0.41
R-88	Castelar	R5	NA	0.00	NA	23.95	11.98
R-100	Castelar	R5	NA	5.81	NA	0.00	2.91
R-83	Mendoza	R5	NA	0.00	NA	17.80	8.90
R-88	Mendoza	R5	NA	0.93	NA	0.82	0.88
R-100	Mendoza	R5	NA	17.26	NA	7.07	12.16
R-101	Castelar	R6	15.13	0.00	NA	0.00	5.04
R-111	Castelar	R6	0.00	0.00	NA	0.00	0.00
R-113	Castelar	R6	2.17	0.87	NA	0.00	1.01
R-101	Mendoza	R6	2.08	0.00	NA	0.00	0.69

R-111	Mendoza	R6	7.00	0.28	NA	0.00	2.42
R-113	Mendoza	R6	2.68	0.00	NA	0.00	0.89
R-41	Castelar	R7	4.18	NA	0.00	NA	2.09
R-57	Castelar	R7	12.37	NA	0.00	NA	6.18
R-93	Castelar	R7	0.00	NA	5.27	NA	2.63
R-41	Mendoza	R7	0.00	NA	0.00	NA	0.00
R-57	Mendoza	R7	0.00	NA	14.58	NA	7.29
R-93	Mendoza	R7	4.65	NA	0.00	NA	2.32
R-70	Castelar	R8	4.94	NA	0.00	0.00	1.65
R-89	Castelar	R8	0.00	NA	0.00	5.97	1.99
R-94	Castelar	R8	0.00	NA	0.00	4.56	1.52
R-70	Mendoza	R8	0.00	NA	0.00	0.00	0.00
R-89	Mendoza	R8	10.23	NA	0.00	11.12	7.12
R-94	Mendoza	R8	0.00	NA	6.48	0.00	2.16
R-96	Castelar	R9	6.50	NA	NA	0.00	3.25
R-102	Castelar	R9	0.00	NA	NA	9.88	4.94
R-96(2)	Castelar	R9	0.00	NA	NA	0.00	0.00
R-96	Mendoza	R9	0.00	NA	NA	17.15	8.58
R-102	Mendoza	R9	0.00	NA	NA	0.00	0.00
R-96(2)	Mendoza	R9	0.00	NA	NA	0.00	0.00

Ensayo de apilamiento de dos loci, F2 cruza (IL9.2.6.1 X IL3.1). Fueron cultivadas 27 líneas recombinantes representativas de las nueve combinaciones recombinantes alélicas posibles para los alelos *S. lycopersicum* y *S. pennelli* de los genes parálogos *VTE3* (1 y 2) evaluando como control sus líneas parentales IL3.1 e IL9.2.6.1 y el Cv. M82 (más detalle en Materiales y Métodos, sección V.2.). En la columna denominada "Planta", comprende las 36 plantas clonales evaluadas. La columna "Ambiente" determina en qué ambiente fue cultivada cada una de las 36 plantas evaluadas. Bajo la columna "Genotipo" se encuentran los 9 genotipos determinados por las combinaciones alélicas para la F2 cruza (IL9.2.6.1 X IL3.1) junto a los parentales IL3.1 e IL9.2.6.1 y el Cv. M82. **A)** Contenidos de tocoferol de tejido de frutos, referido como µg de cada isómero por gramo de pericarpio fresco medidos por HPLC. **B)** Contenidos de tocoferol de tejido de hojas, referido como µg de cada isómero por gramo de tejido fresco medidos por HPLC. **C)** Log₂ Expresión relativa frutos, medidos por qPCR (el cálculo de la misma está explicado en detalle en Materiales y Métodos sección V.4) **D)** Log₂ Expresión relativa hojas, medidos por qPCR (el cálculo de la misma, está explicado en detalle en Materiales y Métodos sección V.4) **E)** Nivel metilación región promotora (McrBc) frutos. **F)** Nivel metilación región promotora (McrBc) hojas. MM⁹ Metilación media del promotor

Tabla S7. Lista de primers utilizados

Nombre ^(a)	Secuencia (5'→3') ^(b)	T _m (°C) ^(c)	Técnica u objetivo ^(d)
MET-VTE3(1)Lyc-II-For	GGGTAAAATTGTAACCTACGCTATGT	55	MCrbc
MET-VTE3(1)Lyc-II-Rev	CTACAAAACCTATTGAAAAAGCCTAA	55	MCrbc
MET-VTE3(1)Penn-II-For	ATGAGATGTGTAGCAATTTAATTAG	55	MCrbc
MET-VTE3(1)Penn-II-Rev	CACACTCAGCTAATTCATCGATTTT	55	MCrbc
Rbcl-For	ATGTCACCACAAACAGAACTAAAGCAAGT	55	MCrbc
Rbcl-Rev	CTTCACAAGCAGCAGCTAGTTCAGGACTCC	55	MCrbc
vte3.2.1For	CAAGGTTGTGGTGGAGTGGT	60	Screening
vte3.2.1Rev	CTCATATGATCATCCTTTGTGATTT	55	Screening
vte3.2.2 Fw	GTGTGTCCTCATAGTTGATGTTCT	58	Screening
vte3.2.2 Rev	CCAAGTTTGCAAAGAGTCAAAA	56	Screening
vte3.1FW	TGGGCTTATCATGGGATGTT	56	Screening
vte3.1 Rev	AAAGACATCTTACTGGGTTCTAGATGA	59	Screening
RT-For-Lyc-VTE3(2)Lean	GACCAACAATGTTTTTGGCATCAT	59.2	q-RT-PCR
RT-Rev-Lyc-VTE3(2)LQ	TTGTGTGATTAGAAAACATGACTGA	57.2	q-RT-PCR
For-RT-VTE3.2-lyc	AGTCTGTTAGTTTTTCAGCCCT	54.3	q-RT-PCR
Rev-RT-VTE3.2-lyc	ACTATTTATTTCAATGAGTTGTGTGA	52	q-RT-PCR
For-RT-VTE3.2-pen	AGCAAGCCTGTGAACCCCTTT	52	q-RT-PCR
Rev-RT-VTE3.2-pen	GGTGGAGAAGCTGTCAAACAGAAATT	54	q-RT-PCR
VTE3 LYC/PENN FOR	TTAGTTGAGACCAAGCTAAGATGTC	60	q-RT-PCR
VTE3 LYC REV	GCGAACCTCTGTGAGAGACC	60	q-RT-PCR
VTE3 PENN FOR	GGTACAAGGATCAGCCAGTAGT	60	q-RT-PCR
RT-VTE3(1)-For	CTTGACCAATCTCCTCATC	60	q-RT-PCR
RT-VTE3(1)-Rev	GCACGCCTTTTCTCCAGG	60	q-RT-PCR
RT-VTE3(2)-For	GCTAAGGCTAGGCAGAAGGAG	60	q-RT-PCR
RT-VTE3(2)-Rev	CAGGCAACCCACCTATGG	60	q-RT-PCR
Metvte3.2For1	GTCTTGGCTGCACTTAGGTATGC	61.8	MCrbc
Metvte3.2Rev1	GCCTCGTATCTCAATGAGCCA	59.9	MCrbc
Metvte3.2For2	TGCAGTCAAACCTCTCTCAAATATGAG	59.9	MCrbc
Metvte3.2For3	TGGCTCATTGAGATACGAGGC	59.9	MCrbc
Metvte3.2Rev3	TTCACCGAAGTGTGAGATTCATAAATATA	59.13	MCrbc
Metvte3.2For4	ATTGGTTGAAGGGTCAACCCATG	61	MCrbc
Metvte3.2Rev4	GGACATTTGCTTGAAGAAGGATAAGC	60.9	MCrbc
Metvte3.2For5	CCTTCCTTTTGGAGATTTTATAGAA	55	MCrbc
Metvte3.2Rev5	TGATAATTGACAAATGTGAGAATTAAT	54	MCrbc
Metvte3.2For6	ACTTCATCCACCTCAGCCTTAC	60	MCrbc
Metvte3.2Rev6	CCACAGAGCTATACGACACCGT	61	MCrbc
Metvte3.2Rev1 bis PEN	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	53	MCrbc

^(a)Nombre con el cual fue denominado el primer, For indica que es un primer *Forward* y Rev que es un primer Reverso. ^(b) indica la secuencia en sentido 5' a 3' para ese primer. ^(c) Indica la temperatura de Anillado del primer o temperatura de *melting* en grados centígrados. ^(d) Indica la técnica o el fin por el cual fue utilizado este primer en el caso del screening de las plantas cruzas IL9-2-6-1xIL3-1 recombinantes.

